A comparison of nominal regression and logistic regression for contingency tables, including the 2 2 2 case in causality

Colignatus, Thomas

Thomas Cool Consultancy Econometrics

19 June 2007

A comparison of nominal regression and logistic regression for contingency tables, including the 2 × 2 × 2 case in causality

Thomas Colignatus, June 19 2007

http://www.dataweb.nl/~cool

Summary

Logistic regression (LR) is one of the most used estimation techniques for nominal data collected in contingency tables, and the question arises how the recently proposed concept of nominal correlation and regression (NCR) relates to it. (1) LR targets the cells in the contingency table while NCR targets only the variables. (2) Where the methods seem to overlap, such as in the 2 × 2 × 2 case, there still is the difference between the use of categories by LR (notably the categories Success, Cause and Confounder) and the use of variables by NCR (notably the variables Effect, Truth and Confounding). (3) Since LR looks for the most parsimonious model, the analysis might be helped by NCR, that is very parsimonious since it uses only the variables and not all the cells of the contingency table. (4) While LR may generate statistically significant regressions, NRC may show that the correlation still is low. (5) Risk difference regression may be a bridge to understand more about the difference between LR and NCR. (6) The use of LR and NCR next to each other may help to focus on the research question and the amount of detail required for it.

Table of contents

1. Introduction

The key point is that logistic regression targets all cells in a contingency table while nominal correlation and regression focus on the variables only. Consider for example a 5 × 7 × 3 × 4 contingency table, thus with 420 cells that may contain all kinds of patterns. The human mind has a psychological craving for clarity and overview but conceptions on what provides that clarity and overview can differ. Both logistic regression (LR) and nominal correlation and regression (NCR) target that same need but their approach is different. In logistic regression we run some 420 regressions (depending on what we take as the success category) and subsequently use statistical tests to determine the most parsimonious model. Nominal correlation and regression focus on the 4 variables only. The latter approach makes a sharper distinction between the collection of the data and the processing of the data. Proper measurement requires that we collect the data in all detail of the 420 cells but decision making might be guided by summary statistics that only concern the 4 variables.

Nominal correlation and regression work since they use the property that a contingency table by itself gives all connections between the variables. Required operations are only

normalization and aggregation, so that, in fact, a contingency table is its own correlation matrix. Subsequently standard statistics takes over and it is possible to determine regression coefficients. When a contingency table shows a correlation between the (assumed) effect variable and some explanatory variable, but theory suggests that the latter would be a confounder, then we don't take that correlation as a causal explanation, and we would collect more data and construct other tables without that correlation to test the theory.

Given that these two methods are so different in their approach, this paper could stop here. Unfortunately, there are the $2 \times ... \times 2$ tables where the variables might be seen as collapsing into the categories, so that these methods might be seen as overlapping. By consequence, this paper has the awkward structure of maintaining the perspective of $n_1 \times n_2 \times ... \times n_k$ contingency tables where the two methods work out differently, while it focusses on the presumed overlap on the 2^n dimensioned tables - which we limit here to the $2 \times 2 \times 2$ case. Working in this way, it seems that we have to understand two universes and the properties of divergent approaches before we can understand the simple issue of the $2 \times 2 \times 2$ case, while we also use that case to understand those two universes. This awkward structure carries the seed of confusion but still has the objective to clarify.

This paper is part of a general project Colignatus (2007e), a work-in-progress book-writing on "Elementary statistics and causality" (ESAC). Insights from epidemiology would be useful for experimental economics but there are differences in conventions and problems of translation. What seems clear-cut might still be a pitfall for someone not trained in the tradition of epidemiology. Colignatus (2007e & f) explain the economic interest in making proper translations. Within this project, Colignatus (2007d) discusses nominal correlation and regression, and Colignatus (2007f) discusses causality in the special $2 \times 2 \times 2$ contingency table with the three variables of (1) an effect, (2) a true cause and (3) a confounder. Within this project, the current paper considers logistic regression.

As Colignatus (2007d) discussed nominal correlation and regression, a natural question is how this relates to logistic regression, one of the most often used techniques for nominal data. This paper gives some first answers. It seems to be one of the paradoxes in empirical practice that while logistic regression leads to regression coefficients and implied measures of association (between categories), it appears difficult to turn these into correlation matrices. Conversely, when Colignatus (2007d) explored the possibility of nominal correlation, the first link was not to logistic regression, but to the volume ratio measure, with the consequence that regression only turned up as a by-product.

Perhaps it would have been faster when Colignatus (2007d) had started with logistic regression anyway, but the original intuition was different and thus the question on the link-up now turns up as a separate issue. Seen conversely, it was fortunate that the discussion did not start out with logistic regression since it allowed the development of the other intuition.

Economics has a strong tradition in aggregation and this may explain the different approach to the contingency table. For example, an economy has millions of products but there still is an aggregate price index. So an economist might be less focussed on the the different categories of a variable but wonder about the variable as a whole. Issues of aggregation are complex, of course, but there are workable solution approaches. Interestingly, a contingency table does not require much theory on aggregation, since the table can be regarded as its own correlation array.

Garson (2007b) mentions two research objectives (for contingency tables): (1) "exploratory", that is targetted at finding "A parsimonious model is the most incomplete model which still achieves a satisfactory level of goodness of fit." (2) "(...) in confirmatory log-linear analysis, one wishes to test a particular model based on theory." It is not entirely clear though where this need for "parsimony" derives from, except from the human psychological need for simplicity. In the past, we did not have computers, so then there might have been a need for numerical simplicity. Nowadays, we would prefer numerical accuracy. When we have done the hard work of collecting all the data, then those actually provide the most accurate description of reality. The probabilities that the data contain (dividing the contingency table by the total) can be seen as parameters, and in a prediction we could merely scale up. If the prediction requires some conditions (e.g. different marginal distributions) then we conditionalize to those particular conditions. We might drop some cells if we don't trust the numbers but this would rather not be for the reason of statistical insignificance but only when we don't trust the data collection process. Hence, it is not entirely clear why we should run all those logistic regressions in order to select "the most parsimonious model" as an objective of itself. On the other hand, there is the human need for clarity and overview, but this psychological need is an objective of an entirely different kind that might be served better by other approaches than logistic regression.

References for logistic regression are in particular Kleinbaum et al. (2003), Garson (2007a, b), Lowry (2007) and Friendly (2007). Theil (1971) gives a very accessible discussion of logit analysis (i.e. using the logarithms of the odds). It turns out that logit analysis is also the most general approach, with possibly more dependent variables, apparently also called "multinomial logistic regression". This carries the implication that

"standard logistic regression" would be "binomial or binary logistic regression" (where one also might assume a Poisson distribution). All this name-tagging is a bit confusing. Apparently computer programs are created for each separate combination, which requires the user to work through the User Guide, but it would be better when the decision tree would be in the computer programme as well. A first step towards clarity is to maintain (i) the distinction between a single equation and systems or equations, (ii) the distinction on the logit transformation or not, and (iii) the distinction on the distribution. Thus instead of only "multinomial" we should also allow for multi-Poisson. A key point is this, though. Logistic regression uses odds, defined on a probability as odds[p] = p / (1 - p). This format leads us to think that only binary possibilities are allowed, thus the binomial model. With $\{p_1, ..., p_n\}$ one would not know how to oppose p_1 to any of the other p_i. However, it is still possible to oppose p_1 to (1 - p_1). Similarly, in the 2 × 2 case with two processes and two probabilities p and q, we can determine an odds ratio odds[p] / odds[q], and at first it seems difficult to generalize this to more dimensions. However, for the $m \times n$ table such odds[$p_{i,j}$] / odds[$p_{k,l}$] easily follow. For more dimensions, we can condition on those. Indeed, the binary cases exist only against the backdrop of aggregation or conditioning over such hidden dimensions. The dichotomous character of the odds thus does not seem to limit the use for more dimensions, yet it will create a tropical forest of figures. The key point is that these extensions retain a dichotomous character in an essential way. Nominal correlation generalizes in a multi-dimensional way. Interestingly, there is some connection between odds ratio and nominal correlation for the 2 × 2 table (if only since that table has only four cells) so one might have the option to see nominal correlation as a multi-dimensional generalization even of the odds ratio.

Thus, there is some scope to think that nominal correlation and regression might provide some general guidance in the analysis of the data. Guidance that might point into the direction of some specific aggregation of the data or going into detail with logistic regression if so required.

In the discussion below, apart from a disclaimer and some preliminaries, we can directly continue with two examples in logistic regression, one that includes a real variable (number of weeks) and one that uses only a contingency table. The two variants help us to understand what logistic regression does. We concentrate on the second example with only a contingency table since it allows the comparison of logistic regression and the suggested measure for nominal correlation and regression.

2. Disclaimer

All disclaimers in Colignatus (2007d, e, f) apply to the following discussion as well, notably that the author is limited in time and other resources.

This paper is part of the project Colignatus (2007e), "Elementary statistics and causality" (ESAC). The present paper is a report on how the author has come to understand issues on correlation and regression in the context of causality. ESAC itself will eventually be written from a didactic point of view, where the current road towards understanding is irrelevant. The following is useful for intermediate documentation and discussion.

In itself, this discussion on regression is a bit of a digression. The focus of this author is on ESAC where the prime problem is to establish causality but not necessarily the size of the impact. The focus got distracted when there appeared to be no clear measure of correlation for contingency tables, so that the idea "correlation isn't causation" had no numerical expression. Devising a "nominal correlation" measure for contingency tables solved that problem. However, this subsequently caused the idea that correlation allows us to also define regression. With hindsight it might have been wiser, in this exploration, to use existing methods of regression and turn those into correlation. Alas, though, there was no original intuition on that, and thus the digression from ESAC took off on its own. The notion of nominal correlation got formulated, and it brought along, by implication, regression coefficients. Which finally caused the question: what do those regression coefficients *mean* ?

Thus, in an ever increasing digression from ESAC, now the question lies on the table how the implied regression coefficients of the suggested correlation measure relate to existing regression methods. This discussion is a bit awkward since we have no strong interest or intuitions on those. Of necessity the discussion will be rather informal and indicative only. We employ the notion that the regression coefficient should give the effect when the variable is increased by one unit.

The basic idea is given by (2007e): True causality should give a correlation coefficient of 1, and if correlation is lower then the difference must be due to confounding, i.e. sooner an error term due to misspecification than to "other causes not mentioned". Built on this base are Colignatus (2007f), i.e. a specific discussion of the $2 \times 2 \times 2$ contingency table in causality, and Colignatus (2007d), i.e. the formulation of a general measure of correlation for contingency tables including also the non-causal ones. Regression comes into the story since it would give the size of the causal impact. Since

(2007d) and (2007f) both refer to regression, it appeared useful to create this separate discussion that they can refer to.

Contingency tables \ Angle	Causality	Correlation
ETC $2 \times 2 \times 2$	(2007 f)	(2007 d)
$n_1 \times ... \times n_k$	Perhaps in the future	(2007 d)

The following discussion contains points that might be useful for Colignatus (2007d) but that paper is already long by itself and is focussed on establishing the consistency of the method of nominal correlation and its interpretation in the geometry of the volume ratio. The reader is referred to Colignatus (2000d) for the explanation of the suggestion for nominal correlation and regression.

Next to the literature that the author has read, there is also some material that he noted on the internet that would be relevant but that has not been digested yet.

3. Preliminaries

3.1 Appendices on epidemiology and economics

Appendix A discusses three research schemes in epidemiology that can be presented in 2×2 tables: the disease-test matrix, the (cohort follow-up) treatment-control matrix and the case-control matrix. Though these have different formats and also different causal assumptions, they can be used for logistic regression on the odds. This discussion also explains why the odds ratio features in the analysis. Originally, this discussion was part of the main body of the text but it appears that the reader is better served with a separate appendix. The matrices there have the 2×2 format so that some conclusions are necessarily limited.

The routine NominalStatistics generates the nominal correlation coefficients and, based upon variance assumptions, also nominal regression coefficients. The transformation on the co-factors is straightforward but new assumptions are on variance, where the variables are regarded as a whole or as an aggregate of their categories. Conceivably, aggregation can takes various formats. Since this paper might be read by non-economists, **Appendix B** contains some examples of aggregation using the Constant Elasticity of Substitution (CES) function. The current implementation of the routine has a modest assumption merely on the variances. It further depends upon the case at hand how one would wish to develop this and to aggregate the variables. When one chooses a

particular form of aggregation then one would determine the appropriate variances and replace the ones currently used in the routine.

It seems that research using logistic regression uses a lot of "static models" in which there is little influence of time. It might be that contingency tables observed over different periods are summed into one table. In economics, the more "dynamic" approach would be to determine the aggregate variables per subperiod and run the test over time. This would be both a test on the stability of the aggregation and the stability of the regression coefficients.

Appendix C contains some contingency tables of Garson (2007b). This appendix merely supplies the nominal correlation and regression matrices, so that one may consider whether they add value to the discussion by Garson.

Appendix D shows that under marginal independence of cause and confounder, the plain vanilla regression of the success on these two marginals generates coefficients that are directly related to the probabilities and risk differences.

3.2 Notation

For notation, whenever possible, we put the data in the *Effect, Truth, Confounding* (ETC222) mold for contingency tables of size 2 × 2 × 2, see Colignatus (2007f). In that mold, the variable *Effect* has two categories or values $\{S, \neg S\}$ with S = Success; the variable *Truth* has two values $\{C, \neg C\}$ with C = Cause; and *Confounding* Z (= C_{ing}) has two values $\{F, \neg F\}$ with F = ConFound.

Standard odds are for a success, Odds[S] = P[S] / P[$\neg S$], yet they can be qualified or conditioned as to the cause and the confounder. Conditional odds are as in Odds[$S \mid C$] = P[$S \mid C$] / P[$\neg S \mid C$]. The Odds Ratio then is Odds[$S \mid C$] / Odds[$S \mid \neg C$], and this ratio would be 1 if C had no impact. The ratio is called "the C odds ratio for success S", comparable to the convention to use "price-elasticity of consumption" for ∂Log[*cons*] / ∂Log[*price*]. Generally, the variable to be explained is Y, the explanatory variable is X and this two-variable Paradise is destroyed by the Znake, the entry of a third entity Z. For the odds ratio Odds[$Y \mid X, Z$] / Odds[$Y \mid X, \neg Z$] the above naming convention gives "the Z odds ratio for the success S conditional on X". Since the success is the obvious target one might also state "the Z odds ratio with the condition X (for the success)". These ratios are called "partial odds ratios" (Garson (2007b)). While this is the general scheme, we will of course meet cases that do not fit the ETC222 mold, for example when there are two causes instead of only one. One can also imagine that when the

numbers of variables and categories rise that this will put a strain on language and one will resort to explicit formulas. It is generally not wise to speak about "the" odds ratio.

The particular example that we consider concerns a case of arthritis with active treatment or a placebo. The treatment is conditioned ("controlled") on sex status, female or male. By happy coincidence the confounder F can also stand for *Female.* It appears to be confusing to deviate from using S for *Success* and to use the symbol S for another variable S (*sex*). So we will avoid this, and the variable for the confounder Z then would denote *sex* ("zex").

3.3 Relation between risk difference, relative risk, odds ratio and nominal correlation

Key concepts are the probability, the risk, risk difference, the relative risk, the odds and the odds ratio. Epidemiology tends to concentrate on disease and the probability of a success becomes a risk for disease or death.

- This is a 2 × 2 table. To allow some generality, we don't mention the meanings of rows and columns.

```
tab = {{a, b}, {c, d}}
```

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

- For a 2 × 2 contingency table, the routines for risk difference and relative risk have the orientation over the rows (deriving from the disease-test study).

```
RiskDiff[tab] // Simplify
```

$$\frac{a}{a+b} - \frac{c}{c+d}$$

```
RRisk[tab] // Simplify
```

$$\frac{a\,(c+d)}{(a+b)\,c}$$

Probabilities and odds can be translated directly, and some distributions might be easier formulated in terms of odds than in terms of probabilities.

- The probability (risk) gives the odds and the odds directly give the probability (risk).

Odds == Odds[Pr]

$$\text{Odds} = \frac{\text{Pr}}{1 - \text{Pr}}$$

Pr == FromOdds[odds]

$$\text{Pr} = \frac{\text{odds}}{\text{odds} + 1}$$

- The (routine for the) odds ratio does not require an orientation.(It was originally defined on the disease-test study but it applies for the case-control study too.)

OddsRatio[{{a, b}, {c, d}}]

$$\frac{a\,d}{b\,c}$$

OddsRatio[{{a, b}, {c, d}} // Transpose]

$$\frac{a\,d}{b\,c}$$

When the probabilities are statistically independent then the risk difference is 0 and both the relative risk and the odds ratio are 1.

tab = PrTable[p, q]

$$\begin{pmatrix} p\,q & p\,(1-q) \\ (1-p)\,q & (1-p)\,(1-q) \end{pmatrix}$$

RiskDiff[tab] // Simplify

0

RRisk[tab] // Simplify

1

OddsRatio[tab]

1

In a logistic regression, the log[odds] or logit would be the variable to be explained.

- LogOdds would be a better term, but Logit has come into use, perhaps to have it standardly clear that natural logs are used, including a reference to the "bit".

  Logit[Pr]

  $$\log\left(\frac{\text{Pr}}{1-\text{Pr}}\right)$$

The log[odds ratio] = log[odds1] - log[odds2] is a difference in logits and will also be called logit too.

The odds and odds ratio have subranges [0, 1] and [1, ∞] while the logit has ranges [-∞, 0] and [0, ∞] which causes some authors to prefer the logit. Similarly, nominal correlation normalizes to [0, 1].

Consider a (regression) equation log[odds[Pr[y]]] == $\alpha + \beta\, x$, for some variable x. Then by implication the log[odds ratio] = $\beta\, \Delta$x and thus β gives the mark up of the log[odds].

- The difference of the log[odds] is the log[odds ratio] (over time or over two dimensions).

  Logit[Pr[t]] – Logit[Pr[t – 1]] == β Δx

  $$\log\left(\frac{\text{Pr}(t)}{1-\text{Pr}(t)}\right) - \log\left(\frac{\text{Pr}(t-1)}{1-\text{Pr}(t-1)}\right) = \beta\,\Delta\text{x}$$

- For small values of β the term e^{β} can be read as $1 + \beta$, and β thus the mark up. For example the 10% mark up.

  E^0.1

  1.10517

For example, when x is dichotomous, then $\Delta x = 1$, and the above then means that odds[t] $\approx (1 + \beta)$ odds[t-1].

For the 2 × 2 case there is a relationship between relative risk and odds ratio on one hand and nominal correlation on the other hand.

- Nominal correlation for 2 × 2 tables is a function of the odds as *bc* * (odds ratio - 1) or may be seen as an overall weighed risk difference, over rows and columns, using *ad* - *bc* = *ad* - *dc* + *dc* - *bc* = *d* (*a* - *c*) - *c* (*b* - *d*). These expressions however are normalized so that they are between -1 and 1.

NominalCorrelation[{{a, b}, {c, d}}] // Simplify

$$\sqrt{\frac{(b\,c - a\,d)^2}{(a+b)\,(a+c)\,(b+d)\,(c+d)}}\ \text{sgn}(a\,d - b\,c)$$

NominalCorrelation[tab] // Simplify

0

It is a bit striking how the odds ratio from the world of gambling is related to the determinant. But people have been dealing with "volume" for ages so it may be that there are intuitions and connections deeply burried in our neural networks. Given this connection between the odds ratio and nominal correlation for the 2 × 2 case, an option is to see nominal correlation as a multi-dimensional odds ratio for multi-way tables. (It doesn't work exactly on Log[*ad / bc*] but it is just an analogy. For a three-way table and *p* the two-way table of the probabilities on success, a Det[*p* / (1 - *p*)] would be exact - but no longer nominal correlation.)

For notation of the 2 × 2 × 2 case, it is useful to mention that these appear in the following manner in *Mathematica*.

- As list

lis = {{{a, b}, {c, d}}, {{e, f}, {g, h}}}

$$\begin{pmatrix} \{a, b\} & \{c, d\} \\ \{e, f\} & \{g, h\} \end{pmatrix}$$

- In TableForm (standardly used in this discussion).

TableForm[lis]

a	*c*
b	*d*
e	*g*
f	*h*

In the ETC222 format, the success will be in the first row (i.e. block of data) and the lack of success in the second row (i.e. block of data), so that the average success probabilities are:

- It is useful to see this since it emphazises that the probabilities must be multiplied by the cell weights. These are not necessarily the same as follows from multiplication of the marginals. Only statistical independence of the summed table allows that approach.

```
p = lis[[1]] / (lis[[1]] + lis[[2]])
```

$$\begin{pmatrix} \frac{a}{a+e} & \frac{b}{b+f} \\ \frac{c}{c+g} & \frac{d}{d+h} \end{pmatrix}$$

- This would test similarity, see **Appendix C**, model A, but not further developed. If zero this is a variation on *a d f g = b c e h,* see also **Appendix D**.

```
Det[p / (1 – p)] // Simplify
```

$$\frac{a\,d}{e\,h} - \frac{b\,c}{f\,g}$$

3.4 Statistical independence

There are two points on statistical independence that may be mentioned here. Colignatus (2007d) contains an example (of R.A. Fisher and smoking) where biological dependence is modelled by statistical independence for subgroups. So it is important to specify what kind of dependence or independence is at issue. The second point is that there is a difference between observing (given) marginals and imposing statistical independence. Much of the mathematics below works already by calculating the specific marginals. But once one starts assuming that only the marginals are sufficient, so that one can substitute any marginal for the specific one, then we are on the track of statistical independence. It is actually a convention to go that road. The discussion below will thus put some emphasis on the notion of statistical independence, and be less strict in that often also a weaker assumption might hold for the specific marginal of the case at hand.

3.5 Relative freedom or conditional independence

A key concept is **relative freedom** or **conditional independence**. For a 2 × 2 × 2 table, when the variable y is explained by x so that we are interested in P[y | x], but where a third variable z may enter as a confounder, then we can eliminate the confounder without problem when y and z are relatively free with respect to x (or conditionally independent given x), denoted as ($y \perp z \mid x$), and meaning that P[y, z | x] = P[y | x] P[z | x]. For the sub-table with value $x = 1$ we should get a relative risk of 1 and for the sub-table of $x = 0$ we should get a relative risk of 1. This same condition of relative freedom is often expressed, though less clear, as P[y | z, x] = P[y | x]. The latter rule however comes in handy for the expression of the odds ratio. PM. In general P[y, z, x] = P[y | z, x] P[z, x] = P[y | z, x] P[z | x] P[x]. Under relative freedom this collapses to P[y, z, x] = P[y | x] P[z | x] P[x]. Thus P[y, z | x] = P[y, z, x] / P[x] = P[y | x] P[z | x]. The deduction works the other way around too.

- Using success S, cause C and confounder F, the following gives the F odds ratio for the condition C, i.e. Odds[S | C, F] / Odds[S | C, ¬F] as named above.

```
Odds[ConditionalPr[S][C, F]] / Odds[ConditionalPr[S][C, ¬ F]]
```

$$\frac{(\text{ConditionalPr}[S][C, F])\,(1 - \text{ConditionalPr}[S][C, \neg F])}{(1 - \text{ConditionalPr}[S][C, F])\,(\text{ConditionalPr}[S][C, \neg F])}$$

```
% /. ConditionalPr[S][C, x_] :→ ConditionalPr[S][C]
```

1

If there is no relative freedom then z can still be a confounder but then we have to deal with its statistical association within the other variables. Depending upon the problem we might want to test whether the relative risks or odds ratios in the subtables have the same deviation from 1. It would make a difference though whether this holds for relative risks or odds ratios. For the causal model it could be important whether something is true for one statistic or the other.

3.6 Some possible tests

According to the scheme by Social Research Methods (1997, 2007), there are (at least) three tests involved:

1. The Cochran-Mantel-Haenszel Test "assumes a common odds ratio" and tests "whether the response is conditionally independent of the explanatory variable

when adjusting for the control variable". This is a bit confusing since this takes P[$S \perp C \mid F$] rather than P[$S \perp F \mid C$].

2. Mantel-Haenszel Test "measures the strength of association by estimating the common odds ratio" or "the average conditional association between the explanatory and the response variable". Thus P[$S \mid C$].
3. Breslow-Day Test tests "homogeneous odds ratio" or "whether the odds ratio between X and Y is the same as in different Z categories".

Christensen (1997) mentions that in the three-way table there are some 8 kinds of directions to consider. All this leaves some questions to answer: (a) The choice of relative risk or odds ratio, for the variable that ought to be constant, may have important consequences for the problem at hand. (b) Relative freedom or conditional independence may be sufficient for causality but not necessary. (c) Even then, relative freedom may be a chance event. (d) A weak relationship might be statistically significant but would still be a weak relationship. Statistical significance in testing is not the same as significance for the problem. (e) The causal model needs consideration, e.g. one cause and a confouder, or two causes.

3.7 Relation to logic

The focus would often not be on statistical independence but on the kind of dependence. For example a logical relationship. A relation like "If it rains then the streets are wet", is an abstraction from observations in a contingency table. The contingency table for a sample of 100 days may look like the below, with third causes for wet streets when it doesn't rain. Those third causes or "error terms" thus show up not in a regression but in the tables themselves.

- A sample of 100 days.

rainmat

Observation count	It rains	It doesn't rain	Total
The streets are wet	25	3	28
The streets are not wet	0	72	72
Total	25	75	100

- Abstracted into a logical relation.

SquareTruthTable["If it rains" ⇒ "The streets are wet"] // Transpose

(If it rains ⇒ The streets are wet)

	If it rains	¬ If it rains
The streets are wet	True	True
¬ The streets are wet	False	True

Such logical relations require "structural zeros" or "fixed zeros". There is often a requirement that tables are "well populated" with the rule of thumb that all entries should be at least 1 while at most 20% would contain less than 5. But this would not hold for such structural zeros. So one would require theory to guide what would be structural zeros while one would indeed need to test whether such zeros are really zero. This might not be a numerical statistical test but a return to the original data collection. A real question would be how figures could be recorded where none should have been.

3.8 The risk difference as a regression coefficient

The following is a "paradigmatic example" of how the risk difference can be seen as a regression coefficient. Here, we return tot the 2 × 2 case, without confounding. For the analysis, this dimension of the matrix can be a bit confounding itself, since instead of considering a regression coefficient for the variable as a whole we now may also consider a regression coefficient for only a value, say the first category.

Let N be the total number of participants in a treatment-control study, T the number actively treated, C the controls (Placebo treatment), E the number with a (positive) effect, such that $E = E_T + E_C$ with the respective subgroups, such that $p_T = E_T / T$ is the treatment cure rate and $p_C = E_C / C$ is the background cure rate, so that $E = p_T T + p_C C$.

TreatmentControlMatrix[Table, 0, Set]

	Effective	Ineffective	Total
Treatment	ET	IT	ET + IT
Controls	EC	IC	EC + IC
Sum	EC + ET	IC + IT	EC + ET + IC + IT

The regression coefficient for an increase of treatment T on the total effect E while allowing the total to change:

$$\Delta E / \Delta T = (p_T (T + \Delta T) + p_C C - E) / \Delta T$$

$$\Delta E / \Delta T = p_T$$

The regression coefficient for an increase of treatment T on the total effect E while keeping the total $T + C$ constant can be found as as the risk difference.

$$\Delta E / \Delta T = (p_T (T + \Delta T) + p_C (C - \Delta T) - E) / \Delta T$$

$$\Delta E / \Delta T = p_T - p_C$$

Note that $T^* = 1 / (p_T - p_C)$ is also called the "number needed to treat", i.e. the number needed to have one single success above the control outcome.

This example provides the idea that we may get a grip on regression coefficients by checking what happens when we add one unit or move one unit. And we would not be surprised if this showed up in some risk difference.

3.9 Saturation

For logistic regression it is standard to start with a "saturated model". We will see that kind of model below. Garson (2007b): "Saturated models always have perfect goodness of fit to the data, but this is a trivial finding. The purpose of log-linear modeling is to eliminate some of the effects while still being able to achieve goodness of fit." The latter need not be true, see our discussion of the "research objectives" above.

Logistic regression allows for as many parameters as there are degrees of freedom so that the regression actually is a more complex manner of averaging. This way of averaging is more elegant since it uses the estimation format, while in the particular case of logistic regression the form of the function warrants that there will be estimates even when the degrees of freedom are zero. As a result of this technique, the correlation coeffficient between y and $\hat{y}$ would always be 1, provided that one includes all combinations, and it must get a value less than 1 if one does not include all combinations - but in less tractical manner how much less.

4. Reproduction of an example of logistic regression (Lowry)

4.1 Data and regression

This example is taken from Lowry (2007): “The following table shows the relationship, for 64 infants, between *X*: gestational age of the infant (in weeks) at the time of birth [column (i)]; and *Y*: whether the infant was breast feeding at the time of release from hospital ["no" coded as "0" and entered in column (ii); "yes" coded as "1" and entered in column (iii)”. The relevant probability per row is the number of score 1 divided by the Sum of score 0 and 1.

The model is no proper contingency table. The variable *X* introduces a cardinal scale. The reason to nevertheless include this example is that it emphasizes that, outside of this example, we are dealing with purely nominal data. It is also useful to observe that the RSquared and T-values that are produced require a particular interpretation, see the next section. In addition, *X,* might be written as *T,* time, so that this might be a dynamic model, and it would be wise to exploit that property if the model were developed further.

- LogOdds is clearer than Logit.

```
TableForm[dat = {{28, 4, 2, 6, 0.3333, 0.5, −0.6931}, {29, 3, 2, 5, 0.4, 0.6667, −0.4055},
    {30, 2, 7, 9, 0.7778, 3.5, 1.2528}, {31, 2, 7, 9, 0.7778, 3.5, 1.2528},
    {32, 4, 16, 20, 0.8, 4., 1.3863}, {33, 1, 14, 15, 0.9333, 14., 2.6391}},
  TableHeadings → {Automatic, {"X", 0, 1, Sum, Pr, Odds, LogOdds}}]
```

	X	0	1	Sum	Pr	Odds	LogOdds
1	28	4	2	6	0.3333	0.5	−0.6931
2	29	3	2	5	0.4	0.6667	−0.4055
3	30	2	7	9	0.7778	3.5	1.2528
4	31	2	7	9	0.7778	3.5	1.2528
5	32	4	16	20	0.8	4.	1.3863
6	33	1	14	15	0.9333	14.	2.6391

```
{X, f0, f1, fsum, p, o, logo} = Transpose[dat];
```

- This reproduces Lowry's estimate.

```
res = Estimate[y == α + λ x, {y → logo, x → X}, {α, λ}, Weights → fsum]
```

$$\Big\{\text{AdjustedRSquared} \to 0.86861,\ \text{BestFitParameters} \to \{\alpha \to -17.2086,\ \lambda \to 0.593403\},$$
$$\text{Correlation} \to 0.945985,\ \text{CovarianceMatrix} \to \begin{pmatrix} 11.1868 & -0.365633 \\ -0.365633 & 0.011988 \end{pmatrix},$$
$$\text{DegreesOfFreedom} \to 4,\ \text{EstimatedVariance} \to 0.209789,$$
$$\text{NumberOfObservations} \to 6,\ \text{ReducedFormQ} \to \text{True},\ \text{RSquared} \to 0.894888,$$
$$\text{StandardDeviation} \to \{3.34466,\ 0.10949\},\ \text{TValues} \to \{-5.14509,\ 5.41972\}\Big\}$$

```
ParametersToE[res]
```

$$\{e^{\alpha} \to 3.36053 \times 10^{-8},\ e^{\lambda} \to 1.81014\}$$

4.2 Interpretation of the results

4.2.1 A standard explanation

In his description of this case, Lowry calls the odds the "odds ratio", expressing that the odds itself are a ratio, but this way of expression differs from the use defined above. Hence we adjust his explanation accordingly.

Lowry (2007): "The exponent of the slope exp(.5934) = 1.81 describes the proportionate rate at which the predicted odds (...) changes with each successive unit of X. In the present example, the predicted odds (...) for X=29 is 1.81 times as large as the one for X=28; (...)"

The parameters for the logodds can be translated back again to the odds and the probabilities.

```
odds[x_] = E^(α + λ x) /. (BestFitParameters /. res)
```

$$e^{0.593403\,x - 17.2086}$$

```
odds[x + 1] / odds[x]  // Simplify
```

1.81014

Note that the average rate of change of the odds is 14 / 0.5 over 5 weeks = 144% if averaged linearly or 195% if done geometrically. Logistic regression comes up with a slightly different number 1.81 or 181% that is directly attributed to the influence of *X*.

- Piecewise average

  Rest[RateOfChange[o]]

  {0.3334, 4.24974, 0., 0.142857, 2.5}

  Average[%]

  1.4452

- Overall geometric.

  (14 / 0.5)^(1/5)

  1.94729

4.2.2 What it means for the probabilities

Once we are used to above explanation then it might suffice. When we are not used to it then it will be helpful to consider the implication for the probabilities.

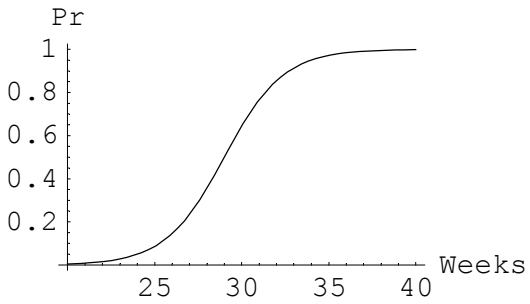
- PM. One might divide by the numerator and then it turns up with a negative power in the denominator, giving an alternative expression $1 / (1 + e^{-(\alpha + \lambda x)})$.

  prob[x_] = FromOdds[odds[x]]

$$\frac{e^{0.593403\,x-17.2086}}{1+e^{0.593403\,x-17.2086}}$$

As in the data, the probability converges to 1.

Plot[prob[x], {x, 20, 40}, AxesLabel → {"Weeks", Pr}];



Thus, actually, the model is that the probability has this logistic shape, from which we derive that the odds ratio would be constant.

```
Pr == Logistic[x, 1, c, λ]
```

$$\text{Pr} = \frac{1}{e^{-x\,\lambda}\,c + 1}$$

```
Odds[Last[%]] // Simplify
```

$$\frac{e^{x\,\lambda}}{c}$$

```
(% /. x → x + 1) / % // Simplify
```

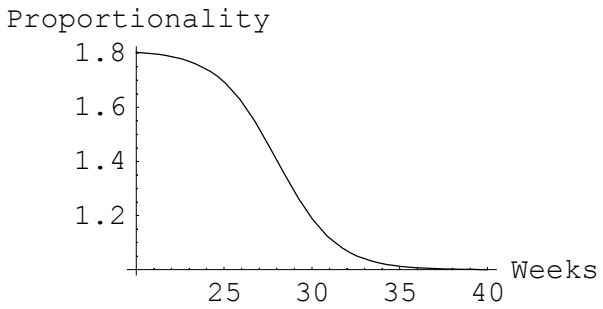
$$e^{\lambda}$$

The probability at some week is the probability of the former week times a proportionality factor. The parameter 1.8 is the proportionality factor at the beginning of the series while over the weeks the proportionality factor drops to 1, as the probability converges to its constant value (in this case 1).

```
prob[x + 1] / prob[x] // Simplify
```

$$\frac{1.81014\,(2.97572 \times 10^{7} + e^{0.593403\,x})}{2.97572 \times 10^{7} + 1.81014\,e^{0.593403\,x}}$$

```
Plot[%, {x, 20, 40}, AxesLabel → {"Weeks", "Proportionality"}];
```



4.2.3 For economists

When we don't use the level of the number of weeks but its logarithm then we get a model structure more familiar to economists. Then ∂ Log[odds] / ∂ Log[weeks] = 18.16 is the week-elasticity of the odds, meaning that if the number of weeks rises by 1% then the odds rise by 18.16%.

```
Estimate[y == γ + η x, {y → logo, x → Log[X]}, {γ, η}, Weights → fsum]
```

$$\left\{\text{AdjustedRSquared} \to 0.870965,\ \text{BestFitParameters} \to \{\gamma \to -61.154,\ \eta \to 18.1616\},\right.$$
$$\text{Correlation} \to 0.94698,\ \text{CovarianceMatrix} \to \begin{pmatrix} 126.913 & -37.1408 \\ -37.1408 & 10.8721 \end{pmatrix},$$
$$\text{DegreesOfFreedom} \to 4,\ \text{EstimatedVariance} \to 0.20539,$$
$$\text{NumberOfObservations} \to 6,\ \text{ReducedFormQ} \to \text{True},\ \text{RSquared} \to 0.896772,$$
$$\left.\text{StandardDeviation} \to \{11.2656,\ 3.29729\},\ \text{TValues} \to \{-5.4284,\ 5.50805\}\right\}$$

Using the log of weeks causes different weights in the regression, thus sleightly different parameters, and also different ways of transformation, but to arrive at the same kind of interpretation. The adjusted R^2 was 86.8% and now is 87.0%. This would still be a logistic model but defined on the logarithm of time. The interpretation is that the odds ratio is not constant. Thus, not transforming to logarithms forces the odds ratios to be constant.

```
Pr == Logistic[Log[x], 1, c, η]
```

$$\text{Pr} = \frac{1}{c\,x^{-\eta} + 1}$$

```
Odds[Last[%]] // Simplify
```

$$\frac{x^{\eta}}{c}$$

```
(% /. x → x + 1) / % // Simplify
```

$$x^{-\eta}\,(x+1)^{\eta}$$

The coefficients 1.810 and 18.16 are a factor 10 apart. To understand this requires us to dig a bit deeper. The first regression equation reads $\log[o] = \alpha + 0.59\ X$, or $o = e^{-17}\ 1.810^{X}$ and the other is $\log[o] = \gamma + 18.16\ \log[X]$ or $o = e^{-61} X^{18.16}$. The differnt kinds of dependency on time X need not be regarded as dramatically different, i.e. for the range of 28 tot 33 weeks used and considering that we will not quickly extend that domain. When we equate the implied relation for X then we find:

$$\gamma + 18.16\ \log[X] = \alpha + \text{Log}[1.1810]\ X$$

```
Solve[-61.154 + 18.16 Log[x] == -17.208 + 0.593 x, Log[x]] // Simplify
```

$$\{\{\log(x) \to 0.0326542\ x + 2.41993\}\}$$

```
Solve[-61.154 + 18.16 "Log[x] " == -17.208 + 0.593 x, x] // Simplify
```

$$\{\{x \to 30.6239\ \text{Log[x]} - 74.1079\}\}$$

We can resolve this by noting that the values of X are around 30 and then determine the power series of Log[*X*] around 30.

```
Series[Log[x], {x, 30, 1}]
```

$$\log(30) + \frac{x-30}{30} + O((x-30)^2)$$

Generally, with η the elasticity and λ the result of logistic regression on the level, and μ the mean of *X,* then

```
γ + η Normal[Series[Log[x], {x, μ, 1}]] == α + λ x // Simplify
```

$$\alpha + \eta + x\,\lambda = \gamma + \eta \log(\mu) + \frac{x\,\eta}{\mu}$$

So that for all practical purposes $\lambda = \eta / \mu$ or $\eta = \mu\ \lambda$. The true relation is that 18.16 / 30 ≈ Log[1.810], where the 30 derives from the approximate mean of *X*.

Another way to describe the model is that the odds in week *X* are given as $o[X] = o[28]\ (1 + \beta)^{X-28}$. Thus the overall average growth over the period o[33] / o[28] = $(1 + \beta)^5$, giving $1 + \beta = 1.95$. Some authors prefer powers of e and then $\lambda = \log[1 + \beta] = 0.667$ and with regression, that weighs the errors along the time path, it becomes 0.59. But, as said, it is also possible to use X^{η}. It may just be what one is accustomed to, and some researchers prefer what they are used to above the size of the error.

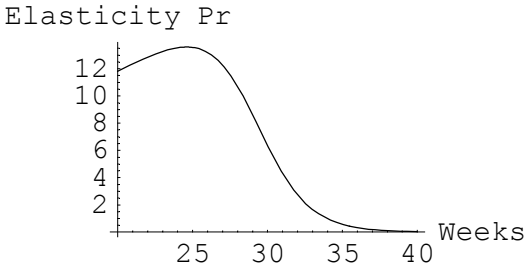
4.3.4 Elasticity of the probability

In economics and epidemiology alike, we are looking for models with constant parameters. Given that probabilities start from 0 and end in 1 it might be difficult to capture this dynamics in a single constant parameter. If the number of weeks rises with 1% then the probability rises with ϵ %, where ϵ is the direct elasticity. In the beginning the elasticity is as high as 1200% but eventually it drops to 0%. This instability in value explains the shift in analysis to the log-odds.

```
Elasticity[prob[x], x] // PowerExpand // Simplify // Rationalize // N
```

$$\frac{1.7658 \times 10^7\ x}{2.97572 \times 10^7 + 2.71828^{0.593403\,x}}$$

```
Plot[%, {x, 20, 40}, AxesLabel → {"Weeks", "Elasticity Pr"}];
```



4.3.5 To proceed

Since this example contains a real-valued variable this example is not a purely nominal model and hence we do not try to relate it to nominal correlation. This is not because such a link-up would be inconceivable but it is only because we did not program that link-up so have nothing to show for now.

5. Reproduction of an example of logistic regression (Friendly)

5.1 The data and their descriptive statistics

5.1.1 The data

This example is taken from Friendly (2007), who refers to Koch & Stokes 1991 for the data. In this case we cannot fully reproduce the numerical estimates. But it concerns a contingency table only so that we can include the nominal statistics. We reorder the data so that they fit the *Effect, Truth, Confounding* (ETC222) mold, see Colignatus (2007f). Normally, one would argue that men and women can have specific risks or effect probabilities, so that this model would be one of two causes and one effect. For our purposes we presume that the sex difference could be a confounder, since it might be that gender is irrelevant.

CT[Set, Default, "Arthritis"]

	Active	Placebo
Some	F 21	13
	M 7	1
None	F 6	19
	M 7	10

CT[Dimensions]

Effect	Some	None
Treatment	Active	Placebo
Sex	F	M

For notation, the variable *Effect E* will have values $\{S, \neg S\}$, *Truth* or *Treatment T* will have values $\{C, \neg C\}$ and *Confounding Z* will have values $\{F, \neg F\}$ or, indeed, $\{F, M\}$. We will employ the word "risk" though this indicates the probability that treatment has some success.

5.1.2 The implied risks, risk differences and relative risks

The data allow the direct calculation of the risk P[$S \mid T, Z$], i.e. probabilities on some effect given treatment status and sex, with $T = \{C, \neg C\}$ en $Z = \{F, \neg F\}$.

- The success probabilities ("Some effect") will feature strongly in the following analysis. Though it reads transposed we better keep this format for the following.

```
dat = CT[Data]; TableForm[p = pmat = dat[[1]] / (dat[[1]] + dat[[2]]) // N,
  TableHeadings → Rest[CT[TableHeadings]]]
```

	F	M
Active	0.777778	0.5
Placebo	0.40625	0.0909091

From these follow the risk differences and the relative risks.

- The risk difference and relative risk for active or passive treatment depend on sex.

```
p[[1]] − p[[2]]
```

{0.371528, 0.409091}

```
p[[1]] / p[[2]]
```

{1.91453, 5.5}

- The risk difference and relative risk for sex status depend on treatment.

```
pt = Transpose[p]; pt[[1]] – pt[[2]]
```

{0.277778, 0.315341}

```
pt[[1]] / pt[[2]]
```

{1.55556, 4.46875}

5.1.3 The odds and odds ratio

The probabilities can also be translated into odds.

- These are the odds.

```
TableForm[o = Odds[p], TableHeadings → Rest[CT[TableHeadings]]]
```

	F	M
Active	3.5	1.
Placebo	0.684211	0.1

The odds ratios are found by dividing over rows and columns, using the formula P[S | X, Z] / P[S | X, ¬Z], for x and Z in alternative combinations of {C, ¬C} and {F, ¬F}. Remember the naming convention of Odds[Y | X, Z] / Odds[Y | X, ¬Z] as "the Z odds ratio with the condition X (for the success)".

1. The active/passive treatment odds ratio for (the condition of) females is 3.5 / 0.68 = 5.1 and for males it is 1 / 0.1 = 10. For females, active treatment has a 5.1 times higher odds than the placebo. For men the odds ratio is twice as high.

```
o[[1]] / o[[2]]
```

{5.11538, 10.}

2. The female/male odds ratio for active treatment is 3.5 / 1 = 3.5 and for the placebo it is 0.68 / 0.1 = 6.8. Under active treatment, females have a 3.5 higher odds than males. Under the placebo females have a 6.8 higher odds. Thus it seems that treatment might have a negative effect on females but actually the true cause is that it has such a strong effect on males.

```
ot = Transpose[o]; ot[[1]] / ot[[2]]
```

{3.5, 6.84211}

5.1.4 The border matrices

The relative risks and odds for the border matrices are weighed averages of the submatrices, with weights that derive from the study and not the population. (See **Appendix A** for the 2 × 2 case.)

- These are the treatment count data, summing out sex.

```
CT[Sum, {"Sex"}, "Arthritis"]
```

	Active	Placebo
Some	28	14
None	13	29

```
RiskDiff[%] // N
```

0.357143

```
RRisk[%%] // N
```

2.15385

```
OddsRatio[%%%] // N
```

4.46154

- These are the sex count data, summing out treatment.

```
CT[Sum, {"Treatment"}, "Arthritis"]
```

	F	M
Some	34	8
None	25	17

```
RiskDiff[%] // N
```

0.214286

```
RRisk[%%] // N
```

1.36

```
OddsRatio[%%%] // N
```

2.89

5.1.5 Average marginal risk or risk difference

The marginals of *Success*, *Cause* and *conFounder* are:

```
{s, c, f} = (First /@ BorderTotals[CT[Data]]) / Add[CT[Data]] // N
```

{0.5, 0.488095, 0.702381}

The weighed average of all probabilities essentially gives the marginal probability of a success. There is a difference with the final outcome of *s* due to the point that *c* and *f* need not be distributed independently while the following assumes that they are. Apparently the (summed) data are close to statistical independence.

```
{c, 1 – c} . p . {f, 1 – f}
```

0.499196

We can write s = P[$S \mid C$] P[C] + P[$S \mid \neg C$] P[$\neg C$] = R c + B (1 - c) = B + (R - B) c, with R the average risk given the cause and B the background risk. Similarly for $s = R_f f + B_f$ (1 - f) for the seeming average risk and seeming backgrond risk. Note that these are *specific* coefficients, with S the dependent variable and C and F the explanatory variables. We can write $\Delta R_{S,C}$ = ΔP[S] / ΔP[C], where the Δ on the right hand side derives from the levels, afterwards divided by N. The values $\Delta R_{S,C} = R - B$ and $\Delta R_{S,F} = R_f - B_f$ are the average risk differences, or in economic terms the average marginal risks. They are regression coefficients when we only consider averages. In general though there will be interaction effects, and c and f need not be distributed independently.

- When f has its observed value.

```
eq1 = S == {C, 1 – C} . p . {f, 1 – f} // Simplify
```

$S = 0.382707\, C + 0.312399$

- The risk difference can be identified directly.

```
Rdif[C] = {1, –1} . p . {f, 1 – f}
```

0.382707

- When *c* has its observed value.

  eq2 = S == {c, 1 − c}.p.{F, 1 − F} // Simplify

  $S = 0.297007\, F + 0.290584$

  Rdif[F] = {c, 1 − c}.p.{1, −1}

  0.297007

- The expression with an explicit interaction effect under independence.

  eq3 = S == {C, 1 − C}.p.{F, 1 − F} // Simplify

  $S = -0.0375631\, F\, C + 0.409091\, C + 0.315341\, F + 0.0909091$

The latter expression has a pitfall. It suggests that *c* and *f* might be distributed independently. But they are not, at least not in general, since the contingency table allows all possible distributions. Thus, the latter expression is a "counterfactual": how the relation would look like, and what the coefficients would be, assuming that there were such independence. PM. Note that the interaction term is small here. The product of *f* c will be small too, so that the overall impact of interaction is small.

See **Appendix D** on what it means that *c* and *f* are marginally independent.

A useful observation is that the above equations also seem (emphasis on *seem*) to imply a value for $\Delta R_{C,F} = \Delta R_{S,F} \,/\, \Delta R_{S,C}$:

- By implication, using the observed value of *c* on the rhs. This is not a proper derivation due to the arbirary mix of *C* and *c, F* and *f.*

  eq4 = {C, 1 − C}.p.{f, 1 − f} == {c, 1 − c}.p.{F, 1 − F} // Simplify

  $1.\, C + 0.0569995 = 0.776067\, F$

5.2 Implied correlation from risk differences

A crucial point to observe is that, alternatively, we could derive c = P[C | S] P[S] + P[C | ¬S] P[¬S] etcetera, reproducing the equations above, but now considering C the effect and S the "cause", and thus doing the regression in the reverse way. When we don't have a model to guide us on the direction of causality we might indeed do the various regressions to see what they imply. Obviously, P[C | S] ≠ 1 / P[S | C], since both conditional probabilities are between 0 and 1. Another point is that $\Delta R_{C,F}$ derived above uses a free C on the left hand side but a given c on the right hand side, which by itself is inconsistent. The proper approach is to actually regard C as the target variable and then derive the proper coefficient.

The general approach is to take the *Effect* category as *S, C* and *F* respectively. Each separate analysis determines the implied probability matrices and recovers the implied equations on the average risk differences. Above we already did this for S but we now repeat this for the other two categories. The result is reproduced in the following. Let us baptise this method as "risk difference regression". (Thus, this uses the categories, and not the variables. We presume that a regression that explains S is the same as one that explains ¬S, and that weights do not cause different coefficients.)

- These equations assume statistical independence. The interaction terms are on the marginals and not on the inner cells.

RiskDiffRegress222[Equations, CT[Data], {S, C, F}]

$$\left\{S = -\frac{119\,F\,C}{3168} + \frac{9\,C}{22} + \frac{111\,F}{352} + \frac{1}{11},\right.$$
$$\left. C = -\frac{291\,S\,F}{3400} - \frac{73\,F}{425} + \frac{63\,S}{136} + \frac{7}{17},\ F = \frac{159\,S\,C}{10556} - \frac{73\,C}{377} + \frac{111\,S}{406} + \frac{19}{29}\right\}$$

Another crucial step is to regard these equations as regressions indeed. When we regard these as regressions then we hypothize that there are correlations between the categories.

cormat = FormalCorrelationMatrix[{S, C, F}] /. Rho[x_, y_] :→ $\rho_{x,y}$

$$\begin{pmatrix} 1 & \rho_{S,C} & \rho_{S,F} \\ \rho_{S,C} & 1 & \rho_{C,F} \\ \rho_{S,F} & \rho_{C,F} & 1 \end{pmatrix}$$

It holds in general that regression coefficients follow from both the correlation matrix and the standard deviations of the categories.

CovarRegress[All, cormat, {σ_S, σ_C, σ_F}]

$$\begin{pmatrix} 0 & -\frac{\sigma_S\,(\rho_{C,F}\,\rho_{S,F}-\rho_{S,C})}{\sigma_C\,(1-\rho_{C,F}^2)} & -\frac{\sigma_S\,(\rho_{C,F}\,\rho_{S,C}-\rho_{S,F})}{\sigma_F\,(1-\rho_{C,F}^2)} \\ -\frac{\sigma_C\,(\rho_{C,F}\,\rho_{S,F}-\rho_{S,C})}{\sigma_S\,(1-\rho_{S,F}^2)} & 0 & -\frac{\sigma_C\,(\rho_{S,C}\,\rho_{S,F}-\rho_{C,F})}{\sigma_F\,(1-\rho_{S,F}^2)} \\ -\frac{\sigma_F\,(\rho_{C,F}\,\rho_{S,C}-\rho_{S,F})}{\sigma_S\,(1-\rho_{S,C}^2)} & -\frac{\sigma_F\,(\rho_{S,C}\,\rho_{S,F}-\rho_{C,F})}{\sigma_C\,(1-\rho_{S,C}^2)} & 0 \end{pmatrix}$$

The correlation matrix and the standard deviations give us six variables. The latter regression matrix gives us six coefficients, that we can identify in the equations. We can solve this when drop the interaction terms. There is some indeterminacy since we have only ratios of the standard deviations. We might also want to impose standard deviations of the marginals as either Binomial or Poisson. The implementation below allows the numerical minimization FindMinimum all freedom and afterwards normalizes the maximal standard deviation to 1 (which need not be a serious value but allows easy comparison with the results).

RiskDiffRegress222[CT[Data], {S, C, F}]

$$\Big\{\text{Equations} \to \Big\{S = -\frac{119\,F\,C}{3168} + \frac{9\,C}{22} + \frac{111\,F}{352} + \frac{1}{11},$$
$$C = -\frac{291\,S\,F}{3400} - \frac{73\,F}{425} + \frac{63\,S}{136} + \frac{7}{17},\ F = \frac{159\,S\,C}{10556} - \frac{73\,C}{377} + \frac{111\,S}{406} + \frac{19}{29}\Big\},$$
$$\text{CovarRegress} \to \begin{pmatrix} 0 & \frac{9}{22} & \frac{111}{352} \\ \frac{63}{136} & 0 & -\frac{73}{425} \\ \frac{111}{406} & -\frac{73}{377} & 0 \end{pmatrix},\ \text{Method} \to \text{Automatic},$$
$$\text{FindMinimum} \to 0.00159022,\ \text{Spread} \to \{0.988278,\ 1.,\ 0.882693\},$$
$$\text{CorrelationMatrix} \to \begin{pmatrix} 1 & 0.407403 & 0.243758 \\ 0.407403 & 1 & -0.0608797 \\ 0.243758 & -0.0608797 & 1 \end{pmatrix}\Big\}$$

In itself this is a nice result. The limitations are: (i) categories and not variables (that in theory might have more than 2 categories), (ii) assumption of independent marginals, (iii) no interaction of the marginals.

This “risk difference regression” thus provides only a limited result, but it allows a stepping stone to understand nominal correlation and regression.

5.3 Plain vanilla regression

We can reorder the data and collect the odds from above. The explanatory variables can be given “dummy” formats.

- This routine assumes the layout of the ETC 2 × 2 × 2 case and then assigns the values of the dummy variables. We have already seen the Pr and Odds above. LogOdds is clearer than Logit.

```
OddsTableETC222[CT[Data]] // N
```

	Cause	Confounder	Success	Sum	Pr	Odds	LogOdds
1	1.	1.	21.	27.	0.777778	3.5	1.25276
2	1.	0.	7.	14.	0.5	1.	0.
3	0.	1.	13.	32.	0.40625	0.684211	−0.37949
4	0.	0.	1.	11.	0.0909091	0.1	−2.30259

- This collects the columns in separate variables.

```
{treat, sex, succ, tot, pvec, o, logo} = Transpose[%];
```

- The numbers of the cause and confounder can be recovered by multiplication of the total with the dummies. If required, the none-effects can be recovered from the total as well.

```
{treat * tot, sex * tot}
```

$$\begin{pmatrix} 27. & 14. & 0. & 0. \\ 27. & 0. & 32. & 0. \end{pmatrix}$$

A plain vanilla linear regression can explain the successes (or the probability of success) from a linear relation of both a general background risk (related to the total number of observations) and the marginal contributions of the two explanatory variables. In this run, we don't include an interaction term; we already saw that it was small, so we get a high R^2.

- The direct explanation of the success in levels.

```
res = Estimate[S == a t + b1 x1 + b2 x2,
    {S → succ, t → tot, x1 → treat * tot, x2 → sex * tot},
    {a, b1, b2}, Weights → tot]
```

$$\Big\{\text{AdjustedRSquared} \to 0.998863,\ \text{BestFitParameters} \to \{a \to 0.114485,\ \text{b1} \to 0.37408,\ \text{b2} \to 0.290808\},\ \text{Correlation} \to 0.99981,\ \text{CovarianceMatrix} \to \begin{pmatrix} 0.000374876 & -0.000118159 & -0.000325739 \\ -0.000118159 & 0.000191104 & 0.0000386867 \\ -0.000325739 & 0.0000386867 & 0.00036423 \end{pmatrix},\ \text{DegreesOfFreedom} \to 1,\ \text{EstimatedVariance} \to 0.0956767,\ \text{NumberOfObservations} \to 4,\ \text{ReducedFormQ} \to \text{True},\ \text{RSquared} \to 0.999621,\ \text{StandardDeviation} \to \{0.0193617, 0.013824, 0.0190848\},\ \text{TValues} \to \{5.91294, 27.0601, 15.2377\}\Big\}$$

When we also include an "interaction term" for the joint occurrence of treatment and sex then we get a 100% explanation, though the statistics become undefined. The regression now reproduces our earlier allocation of the impacts. The regression routine actually is only used as a different manner to calculate the averages.

- We have seen these coefficients above. It doesn't matter if we use levels or rates; this explains the probabilities.

```
res =
  Estimate[S == a + b1 x1 + b2 x2 + b3 joint, {S → pvec, x1 → treat, x2 → sex,
     joint → treat * sex}, {a, b1, b2, b3}, Weights → tot]
```

$$\Big\{\text{AdjustedRSquared} \to \text{ComplexInfinity},\ \text{BestFitParameters} \to \{a \to 0.0909091,\ \text{b1} \to 0.409091,\ \text{b2} \to 0.315341,\ \text{b3} \to -0.0375631\},\ \text{Correlation} \to 1.,\ \text{CovarianceMatrix} \to \begin{pmatrix} \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} \\ \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} \\ \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} \\ \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} & \text{ComplexInfinity} \end{pmatrix},\ \text{DegreesOfFreedom} \to 0,\ \text{EstimatedVariance} \to \text{ComplexInfinity},\ \text{NumberOfObservations} \to 4,\ \text{ReducedFormQ} \to \text{True},\ \text{RSquared} \to 1.,\ \text{StandardDeviation} \to \{\text{ComplexInfinity}, \text{ComplexInfinity}, \text{ComplexInfinity}, \text{ComplexInfinity}\},\ \text{TValues} \to \{0, 0, 0, 0\}\Big\}$$

The probabilities thus are functions of the dummy values.

```
prob[x1_, x2_] = (a + b1 x1 + b2 x2 + b3 x1 x2) /. (BestFitParameters /. res)
```

$$-0.0375631\ x2\ x1 + 0.409091\ x1 + 0.315341\ x2 + 0.0909091$$

```
TableForm[pest = Outer[prob, {1, 0}, {1, 0}],
  TableHeadings → Rest[CT[TableHeadings]]]
```

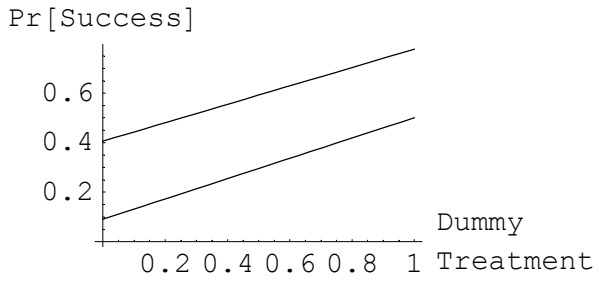
	F	M
Active	0.777778	0.5
Placebo	0.40625	0.0909091

As remarked, we already saw these coefficients above. It so happens, which just happens in this data example, that the marginals c and f are close to independence, and in that case the coefficients of $S == \{C, 1 - C\} \,.\, p \,.\, \{F, 1 - F\}$ are close to those generated by the plain vanilla regression. **Appendix D** explains. The reason is that the error is zero so that the coefficients must be equal too. Alternatively, the marginals are not independent, and then the interaction term also captures the error $S - \{C, 1 - C\} \,.\, p \,.\, \{F, 1 - F\}$.

The regression format also shows its limited value. The dummies and their combinations cover all the data cells and thus all cells are provided with a 100% explanation. For the joint effect we might also have plugged in the salary of the US President and found a coefficient. The only reason not to plug in the salary of the US President is that it makes no sense to do so. But realizing this, helps to be sober about what this kind of regression achieves.

Nevertheless, the assumption of linearity allows interpolation between the 0 and 1 extremes, e.g. on the size of the dose.

```
Plot[{prob[x, 1], prob[x, 0]}, {x, 0, 1},
    AxesLabel → {"Dummy\nTreatment", "Pr[Success]"}];
```

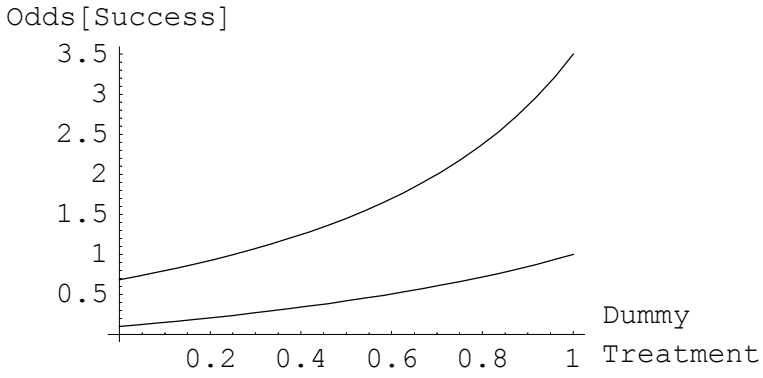


And this is the transform of the odds.

```
odds[x1_, x2_] =  Odds[prob[x1, x2]]
```

$$\frac{-0.0375631\ x2\ x1 + 0.409091\ x1 + 0.315341\ x2 + 0.0909091}{0.0375631\ x2\ x1 - 0.409091\ x1 - 0.315341\ x2 + 0.909091}$$

```
Plot[{odds[x, 1], odds[x, 0]}, {x, 0, 1},
    AxesLabel → {"Dummy\nTreatment", "Odds[Success]"}];
```



In the Lowry (2007) example, the plain vanilla regression was less attractive since an extrapolation of the number of weeks might result into a forecast with probabilities outside the range [0, 1]. For a pure contingency table it is less likely that such out-of-range results arise since all input variables are themselves in that domain. Nevertheless, relationships might be nonlinear, or the log-odds might be linear, so we now consider logistic regression.

In these plain vanilla regression we worked in the direction from no-interaction to interaction. For logistic regression it is more informative to first use full interaction and then remove it.

5.4 Logistic regression

5.4.1 Logistic regression with an interaction effect

We can reproduce the earlier findings by using logistic regression that includes an "interaction effect". By consequence the degrees of freedom are zero and the regression produces some seemingly troubling statistics. These are not to worry about, as will become clear shortly.

- We keep the order of cause and confounder (which differs from Friendly (2007)).

```
res = Estimate[y == a + b1 x1 + b2 x2 + b3 x1 x2,
    {y → logo, x1 → treat, x2 → sex}, {a, b1, b2, b3}, Weights → tot]
```

```
{AdjustedRSquared → ComplexInfinity, BestFitParameters →
   {a → -2.30259, b1 → 2.30259, b2 → 1.9231, b3 → -0.670333}, Correlation → 1.,
 CovarianceMatrix → {{ComplexInfinity, ComplexInfinity, ComplexInfinity, ComplexInfinity},
   {ComplexInfinity, ComplexInfinity, ComplexInfinity, ComplexInfinity},
   {ComplexInfinity, ComplexInfinity, ComplexInfinity, ComplexInfinity},
   {ComplexInfinity, ComplexInfinity, ComplexInfinity, ComplexInfinity}},
 DegreesOfFreedom → 0, EstimatedVariance → ComplexInfinity,
 NumberOfObservations → 4, ReducedFormQ → True, RSquared → 1.,
 StandardDeviation → {ComplexInfinity, ComplexInfinity, ComplexInfinity, ComplexInfinity},
 TValues → {0, 0, 0, 0}}
```

When we translate the estimated coefficients then we recognize the odds ratios 10, 6.8 and 0.51 (times 10) that we calculated earlier.

```
ParametersToE[res]
```

$\{e^{a} \to 0.1, e^{b1} \to 10., e^{b2} \to 6.84211, e^{b3} \to 0.511538\}$

```
odds[x1_, x2_] = E^(a + b1 x1 + b2 x2 + b3 x1 x2) /. (BestFitParameters /. res)
```

$e^{-0.670333\, x2\, x1+2.30259\, x1+1.9231\, x2-2.30259}$

We might make a plot but that is less useful given the dichotomy. The four combinations are the following.

- These are exactly the odds as we already calculated above.

```
TableForm[oest = Outer[odds, {1, 0}, {1, 0}],
  TableHeadings → Rest[CT[TableHeadings]]]
```

	F	M
Active	3.5	1.
Placebo	0.684211	0.1

- The probabilities follow from there.

```
prob[x1_, x2_] = FromOdds[odds[x1, x2]]
```

$$\frac{e^{-0.670333\, x2\, x1+2.30259\, x1+1.9231\, x2-2.30259}}{1+e^{-0.670333\, x2\, x1+2.30259\, x1+1.9231\, x2-2.30259}}$$

The four combinations for the probabilities of success are the following.

- These are exactly the probabilities as we already calculated above.

```
TableForm[pest = {{aa, bb}, {cc, dd}} = Outer[prob, {1, 0}, {1, 0}],
  TableHeadings → Rest[CT[TableHeadings]]]
```

	F	M
Active	0.777778	0.5
Placebo	0.40625	0.0909091

And the odds ratios are also the same.

- Odds ratios for treatment per sex. We can use the odds table directly or the individual probabilities (where the latter may be instructive if you have never seen this before).

```
oest[[1]] / oest[[2]]
```

```
{5.11538, 10.}
```

```
{aa / (1 – aa) / (cc / (1 – cc)), (bb / (1 – bb)) / (dd / (1 – dd))}
```

```
{5.11538, 10.}
```

- Odds ratios for sex per treatment.

```
ot = Transpose[oest]; ot[[1]] / ot[[2]]
```

```
{3.5, 6.84211}
```

```
{aa / (1 – aa) / (bb / (1 – bb)), cc / (1 – cc) / (dd / (1 – dd))}
```

```
{3.5, 6.84211}
```

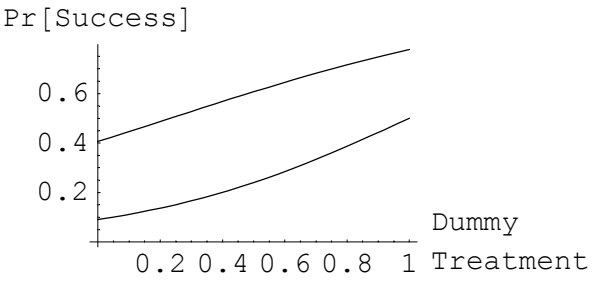
Hence:

- Logistic regression is an efficient technique to determine the odds and odds ratios that are contained in a contingency table. These odds ratios can also be determined directly, but that requires us to calculate various conditional probabilities, in various combinations. This is easily done in *Mathematica*, but for most programs the technique of logistic regression allows a gain in handling-efficiency by using dummies.
- An advantage of logistic regression seems also the easy link up to probabilities and other transforms - though this may be a matter of programming too. The

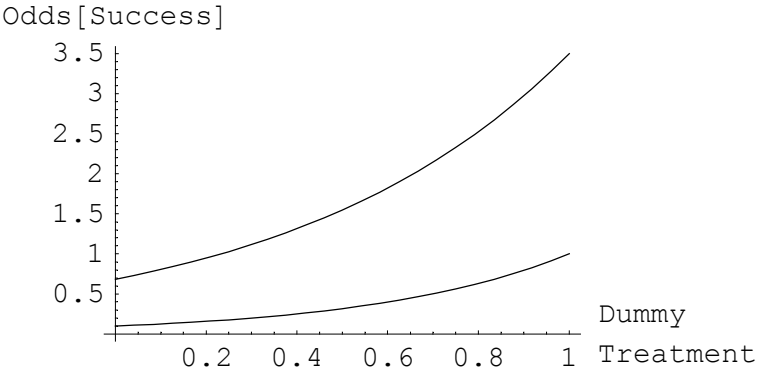
possibility to include real variables is a clear advantage (see the Lowry (2007) example).

- When the dichotomy conceivable would allow intermediate values, e.g. when the dichotomy concerns a low and a high dose, then one might interpolate to get a forecast or expected value.

```
Plot[{prob[x, 1], prob[x, 0]}, {x, 0, 1},
    AxesLabel → {"Dummy\nTreatment", "Pr[Success]"}];
```




```
Plot[{odds[x, 1], odds[x, 0]}, {x, 0, 1},
    AxesLabel → {"Dummy\nTreatment", "Odds[Success]"}];
```



- Logistic regression reproduces all odds when we include all "interaction effects" - and then we should not be distracted by estimation statistics that are indeterminate or ComplexInfinity.
- Logistic regression allows simulation of counterfactuals, by deliberately dropping some "interaction effect". This amounts to averaging out certain combinations. In that case, the degrees of freedom rise and we get determinate

estimation statistics. The values of the correlation coefficient and the T-values should be read in a reverse direction: when the correlation stays close to 1 then the error in dropping an "interaction effect" need not be too large. This is the topic of the next subsection.

5.4.2 Logistic regression without an interaction effect

In this estimation we drop the "interaction effect". Now the estimation output has a normal "look and feel". The estimation output should however be read in a reverse manner. That the correlation remains close to 1 means that there is a small error in averaging out that interaction.

- The dummies are independent and have zero covariance. The dummies also have a negative correlation to the constant.

```
res = Estimate[y == a + b1 x1 + b2 x2,
    {y → logo, x1 → treat, x2 → sex}, {a, b1, b2}, Weights → tot]
```

$$\Big\{\text{AdjustedRSquared} \to 0.945677,\ \text{BestFitParameters} \to \{a \to -2.03835,\ \text{b1} \to 1.83074,\ \text{b2} \to 1.56803\},\ \text{Correlation} \to 0.990905,\ \text{CovarianceMatrix} \to \begin{pmatrix} 0.0995725 & -0.0663816 & -0.0663816 \\ -0.0663816 & 0.132763 & 0. \\ -0.0663816 & 0. & 0.132763 \end{pmatrix},\ \text{DegreesOfFreedom} \to 1,\ \text{EstimatedVariance} \to 0.132763,\ \text{NumberOfObservations} \to 4,\ \text{ReducedFormQ} \to \text{True},\ \text{RSquared} \to 0.981892,\ \text{StandardDeviation} \to \{0.315551, 0.364367, 0.364367\},\ \text{TValues} \to \{-6.45965, 5.02443, 4.30343\}\Big\}$$

Averaging out the interaction appears to have little effect on the cause but most on sex differences.

```
ParametersToE[res]
```

$$\{e^{a} \to 0.130244,\ e^{\text{b1}} \to 6.23847,\ e^{\text{b2}} \to 4.79718\}$$

The treatment odds ratio is 6.2 now and lies between the 5.1 and 10 that we had before. The female/male odds ratio for active treatment now is 4.8 and lies between the values 3.5 and 6.8 that we observed earlier when we kept account of all combinations. It will be useful to show what this average means.

```
odds[x1_, x2_] = E^(a + b1 x1 + b2 x2) /. (BestFitParameters /. res)
```

$$e^{1.83074\,x1+1.56803\,x2-2.03835}$$

- The estimated odds now differ from the true values above.

```
TableForm[oest = Outer[odds, {1, 0}, {1, 0}],
  TableHeadings → Rest[CT[TableHeadings]]]
```

	F	M
Active	3.89781	0.812521
Placebo	0.624802	0.130244

```
prob2[x1_, x2_] = FromOdds[odds[x1, x2]]
```

$$\frac{e^{1.83074\,x1+1.56803\,x2-2.03835}}{1+e^{1.83074\,x1+1.56803\,x2-2.03835}}$$

It is a matter of discussion of course what the "true values" are. The observed contingency table is only a random draw from the universe of this arthritis problem. Perhaps the true model is that there are no interaction effects. In that case the following would be the true probabilities.

- The estimated probabilities now differ from the true values above.

```
TableForm[pest = Outer[prob2, {1, 0}, {1, 0}],
  TableHeadings → Rest[CT[TableHeadings]]]
```

	F	M
Active	0.795827	0.448282
Placebo	0.38454	0.115235

Above estimation imposed the condition that the odds ratios would be the same over the columns or rows, respectively.

- The treatment effect (active versus passive) is independent of sex.

```
oest[[1]] / oest[[2]]
```

{6.23847, 6.23847}

- The female/male odds ratio now holds irrespective whether the treatment is active or placebo.

```
ot = Transpose[oest]; ot[[1]] / ot[[2]]
```

{4.79718, 4.79718}

There doesn't seem to be an easy expression how this averaging takes place, other than by the format of logistic regression itself. We could work out the equality condition and insert it in the "plain vanilla" regression - so that it would no longer be plain vanilla; and

since it is imposed in logistic regression, we would be doing logistic regression in a weird way. What can be useful to observe is that these odds ratios differ from the odds ratios of the border matrices that arise by summing out the other variable. (This would be obvious in itself, since the odds ratios with the main effects only (no interaction) refer to three variables and not two; yet, given the complexity at interpretation, what is obvious may sometimes be obscured again.)

What is important on above condition, though, is not quite that the odds ratios are the same but that the condition does not concern any equality of the risk differences or the relative risks. These can still be all over the place.

- The risk difference and relative risk for treatment status still depends on sex.

  pest[[1]] – pest[[2]]

  {0.411287, 0.333047}

  pest[[1]] / pest[[2]]

  {2.06955, 3.89016}

- The risk difference and relative risk for sex status still depends on treatment.

  pt[[1]] – pt[[2]]

  {0.347545, 0.269305}

  pt = Transpose[pest]; pt[[1]] / pt[[2]]

  {1.77528, 3.33701}

It is not by itself obvious why it would be important that the odds ratios would be the same. For the case considered it might be more important that the relative risks are kept the same. One would use another estimation format then.

5.5 Intermediate conclusions

- In summary, logistic regression is an efficient way to allocate average effects. The logistic format may produce results in a robust fashion. If one allows for all interactive terms then the degrees of freedom drop to zero and regression produces results with seemingly troubling statistics. This is nothing to worry about since that regression actually produces the true situation. When we would include all interaction effects then we would be be no further than section 5.2 on the implications of the risk differences. When we would exclude some interactions then the earlier question on our research objective becomes acute: why would we do so ?
- The correlation matrices produced in a logistic regression (implied by the covariance matrix) cannot be used to say anything about the correlation of the variables. We want the correlations in the true situation, but then the numbers produced are indeterminate.
- The series of logistic regressions done above all regard *S* as the effect variable. We would have to run the same schemes with *C* and *F* as effect variables to get the whole picture. We might even do this simulataneously.

6. ETC analysis

We now apply the *Effect, Truth and Confounding* analysis of Colignatus (2007f). This analysis is targetted on (relative) risk and not on the odds.

The ETC format regards four specific conditional probabilities *r, b, w, v* as the key parameters and derives the total cell results from the marginal distributions *c* and *f* and an interaction parameter *q,* while the whole table adds up to 1 (to scale up with *N*).

```
Clear[c, f]; lis = SafetyToETCArray[{c, r, b}, {f, q}, {w, v}];

TableForm[lis, TableHeadings → CT["ETC", TableHeadings]]
```

		Cause	¬ Cause
Success	Confounder	$(c-(1-f)\,q)\,(1-w)$	$(-c+f\,(1-q)+q)\,(1-v)$
	¬ Confounder	$(1-f)\,q\,r$	$b\,(1-f)\,(1-q)$
¬ Success	Confounder	$(c-(1-f)\,q)\,w$	$(-c+f\,(1-q)+q)\,v$
	¬ Confounder	$(1-f)\,q\,(1-r)$	$(1-b)\,(1-f)\,(1-q)$

Applying this analysis to the arthritis case gives these conclusions:

- The probabilities that we calculated above show up as the Risk probabilities (namely the risk of some recovery).
- This is not a “simple” causal scheme. (A simple cause means: an effect iff a cause.)
- There is no relative freedom i.e. no conditional independence.
- There is no Simpson paradox.
- There is a strong difference between the true relative risk of 5.5 when the confounder Female is absent, and the relative risk 1.9 when it is present. Given that the confounder Female has a prevalence of 70% it dominates the overall outcome. The confounder reduces safety importantly, suggesting that it has a real effect. It may well be that the ETC model does not apply, so that the true situation is one of sex-specific risks (cure rates).

- The risk is having some positive treatment effect.

```
((res = ETCStatistics[CT[Data]] // N) // MatrixForm)
```

Matrix ETCStatistics["Cause, True, Ratio"]

	Cause	¬ Cause	Total
Success	0.28	0.04	0.32
¬ Success	0.28	0.4	0.68
Sum	0.56	0.44	1.

Matrix ETCStatistics["Cause"]

	Cause	¬ Cause	Total
Success	28	14	42
¬ Success	13	29	42
Sum	41	43	84

Matrix ETCStatistics["Confounder"]

	Cause	¬ Cause	Total
Confounder	27	32	59
¬ Confounder	14	11	25
Sum	41	43	84

Matrix ETCStatistics["Seeming"]

	Confounder	¬ Confounder	Total
Success	34	8	42
¬ Success	25	17	42
Sum	59	25	84

$$\left(\begin{array}{l}
N \to 84. \\
\text{NSuccess} \to 42. \\
\text{NCause} \to 41. \\
\text{NConfounder} \to 59. \\
\text{MarginalPr(Success)} \to 0.5 \\
\text{MarginalPr(Cause)} \to 0.488095 \\
\text{MarginalPr(Confounder)} \to 0.702381 \\
\text{IndependentPr(Truth, Confounding)} \to \text{False} \\
(\text{Success} \perp \neg\,\text{Confounder})(\text{Cause}) \to \text{False} \\
(\text{Success} \perp \neg\,\text{Confounder})(\neg\,\text{Cause}) \to \text{False} \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Cause}, \neg\,\text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Cause}, \neg\,\text{Confounder}] \to 0.0909091 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Cause}, \text{Confounder}] \to 0.777778 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Cause}, \text{Confounder}] \to 0.40625 \\
\text{Risk} \to \begin{pmatrix} 0.777778 & 0.40625 \\ 0.5 & 0.0909091 \end{pmatrix} \\
\text{Interaction} \to \{\text{Add} \to -0.0375631, \text{Times} \to -3.58547\} \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Cause}] \to 0.682927 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Cause}] \to 0.325581 \\
\text{ConditionalPr}[\,\text{Cause}\,][\,\text{Confounder}] \to 0.457627 \\
\text{ConditionalPr}[\,\text{Cause}\,][\,\neg\,\text{Confounder}] \to 0.56 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Confounder}] \to 0.576271 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Confounder}] \to 0.32 \\
\text{RRisk(True)} \to 5.5 \\
\text{RRisk(Cause)} \to 2.09756 \\
\text{RelativePr(Confounder)} \to 0.817191 \\
\text{RRisk(Seeming)} \to 1.80085 \\
\text{ETCAdjustedRRisk} \to \{2.09756, 5.5, 1.91453, 2.98163\} \\
\text{Conditions} \to \{\text{True}, \text{True}, \text{False}, \text{True}, \text{False}\} \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\text{Cause}, \neg\,\text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\neg\,\text{Cause}, \neg\,\text{Confounder}] \to 0.909091 \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\text{Cause}, \text{Confounder}] \to 0.222222 \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\neg\,\text{Cause}, \text{Confounder}] \to 0.59375 \\
\text{Safety} \to \begin{pmatrix} 0.222222 & 0.59375 \\ 0.5 & 0.909091 \end{pmatrix} \\
\text{SimpleCauseQ} \to \begin{pmatrix} \text{False} & \text{False} \\ \text{False} & \text{False} \end{pmatrix} \\
\text{ETCSimpson} \to \{\text{Necessary} \to \text{False}, \text{Sufficient} \to \{\text{True}, \text{True}, \text{False}\}\}
\end{array}\right)$$

7. Nominal regression

7.1 Nominal statistics of the arthritis case

For nominal correlation we think in terms of variables and not categories. The correlation and regression coefficients hold for *E* and *T* and not for *S* and *C*. This may be confusing for the 2 × 2 case but makes sense for the $n_1 \times n_2 \times ... \times n_k$ case. In the above we have been guided into thinking in terms of the marginal probabilities with values between 0 and 1 but for nominal correlation it is more adequate to think in the levels, with the interpretation of a regression coefficient as the effect of adding or moving a unit.

resns = NominalStatistics[CT[Data]]

$$\Big\{\text{ContingencyTableQ} \to \text{True},\ \text{OverallCorrelation} \to 0.542123,$$
$$\text{Length} \to \{2, 2, 2\},\ \text{EffectiveNumberOfCategories} \to \{2., 1.99887, 1.71846\},$$
$$\text{Variance} \to \{1., 0.999433, 0.836168\},\ \text{BorderTotals} \to \begin{pmatrix} 42 & 42 \\ 41 & 43 \\ 59 & 25 \end{pmatrix},$$
$$\text{BorderMatrices} \to \Big\{\{1, 2\} \to \begin{pmatrix} 28 & 14 \\ 13 & 29 \end{pmatrix},\ \{1, 3\} \to \begin{pmatrix} 34 & 8 \\ 25 & 17 \end{pmatrix},\ \{2, 3\} \to \begin{pmatrix} 27 & 14 \\ 32 & 11 \end{pmatrix}\Big\},$$
$$\text{NominalCorrelationMatrix} \to \begin{pmatrix} 1. & 0.392654 & 0.288471 \\ 0.392654 & 1. & -0.198373 \\ 0.288471 & -0.198373 & 1. \end{pmatrix},$$
$$\text{CovarMat} \to \begin{pmatrix} 1. & 0.392543 & 0.263785 \\ 0.392543 & 0.999433 & -0.181345 \\ 0.263785 & -0.181345 & 0.836168 \end{pmatrix},$$
$$\text{CovarRegress} \to \begin{pmatrix} 0. & 0.468441 & 0.417062 \\ 0.490575 & 0. & -0.371637 \\ 0.396077 & -0.337013 & 0. \end{pmatrix}\Big\}$$

7.2 Comparison of the covariance matrices of the two types of regression

In logistic regression, the full regression is indeterminate, so we may consider the main effects covariance matrix and compare it to the covariance matrix of nominal regression. The comparison verifies that the covariance matrices have different structures: (i) logistic regression has the dummies, uses their categories, and only considers the direction towards the success, (ii) nominal correlation has all three variables and all their pairwise correlations, using all categories per variable.

7.3 Coefficients for variables and not categories

7.3.1 The total matrix and its meaning

For the matrix C and variables $x = \{x_1, x_2, x_3\}$, the relation is $x = C.x + \epsilon.$. The variable that is explained has coefficient 0 on the diagonal. The matrix C of regression coefficients ("CovarRegress") is not symmetric since it matters in regression what the explained variable is.

- This is the nominal regression matrix, selected from above output. The entries are the variables and not the categories.

NominalStatistics[Results, CovarRegress, CT[Variables]]

	Effect	Treatment	Sex
Effect	0.	0.468441	0.417062
Treatment	0.490575	0.	−0.371637
Sex	0.396077	−0.337013	0.

The variable to be explained is the effect, so we consider the first row in the nominal regression and see that treatment and sex have a positive contribution. Regression coefficients can be interpreted by the effect of adding a unit (allow the total to be raised) or by moving a unit (keeping the same total). The order of presentation determines the direction or sign.

Though we now consider variables and not categories, it is instructive to link up to section 5.2, where we imputed correlation from the risk differences.

The interpretation is:

- Treatment and Effect have a correlation coefficient of 39.3%. This compares to the imputed correlation between success and active treatment of 40%. (It is actually closer to the risk difference of 0.382707.)
- Sex and Effect have a correlation coefficient of 28.8%. This compares to the imputed correlation between success and female sex of 24%. (It is actually closer to the risk difference of 0.297007.)
- Treatment and Sex have a negative correlation of -19.8%. This differs from the imputed correlation between active treatment and female sex of -6%. Nominal correlation keeps account of the differences in the submatrices, and these are more different for these variables.

- The latter difference is related to the difference in regression coefficients. These are higher for nominal correlation since the latter coefficient appears as in the denominator as $(1 - \rho^2_{T,Z})$ respectively $(1 - \rho^2_{C,F})$.
- Moving 1 unit from Active to Placebo would be like moving 0.468 units from Some to None. Or conversely, moving 1 unit from Placebo to Active would be like moving 0.468 units from None to Some.
- Moving 1 unit from Female to Male would be like moving 0.417 units from Some to None. Or conversely, moving 1 unit from Male to Female would be like moving 0.417 units from None to Some.
- In the other rows: The size of the impact depends on what we take as the variable that is explained.

Differences in values compared to the pure average risk difference approach arise because the philosophy behind nominal correlation is different, with the possibility to quickly extend into more dimensions (categories and variables). But that philosophy still has more to do with unit changes and risk differences than with odds.

7.3.2 The first regression line

Though the regression coefficients have already been calculated, it is useful to redo some steps to show how that has been done.

- These are the standard deviations.

```
stdev = Variance /. resns // Sqrt
```

```
{1., 0.999717, 0.914422}
```

- This does the regression for the first variable from a correlation matrix and the standard deviations. This allows us to identify the impact of the standard deviations.

```
CovarRegress[NominalCorrelationMatrix /. resns, {σS, σT, σZ}]
```

$$\left\{\frac{0.468308\,\sigma_S}{\sigma_T},\ \frac{0.381371\,\sigma_S}{\sigma_Z}\right\}$$

- These are the two regression coefficients for treatment and sex.

```
coefs = % /. Thread[{σS, σT, σZ} → stdev] // Simplify
```

```
{0.468441, 0.417062}
```

The latter are close to the overall average effect 0.5, which makes sense since *c* and *f* are close to independence. Note that these coefficients are for the variables and not for the values (categories) of them.

7.3.3 Adding a unit

As said, regression coefficients derive their interpretation from adding or moving a unit. Understanding is best facilitated by thinking in terms of levels and where useful translate to marginal probabilities.

- If *c* and *f* would weigh in 100% then we should allow for an error, currently about 4%. The same model would apply for using values {1, 0} for the variables.

```
s == coefs.{c, f} + Error[]
```

$$0.5 = \text{Error}() + 0.52158$$

```
s == coefs.{1 c + 0 (1 - c), 1 f + 0 (1 - f)} + Error[]
```

$$0.5 = \text{Error}() + 0.52158$$

In general, though, the variables are not well represented by a single value, since in more-dimensional tables there can be lists of categories, and then the assignment of {1, 0} also breaks down. One would generally use an aggregator (possibly akin to a variance measure). It depends upon the case at hand what aggregator works best over time. Though the variables would have their own aggregator the following example assumes a uniform aggregator that applies to all variables.

- This would be the way to understand it.

```
eqn = (func[{s, 1 - s}] == coefs.{func[{c, 1 - c}], func[{f, 1 - f}]} + Error[])
```

$$\text{func}(\{0.5, 0.5\}) = \text{Error}() + 0.468441\ \text{func}(\{0.488095, 0.511905\}) + 0.417062\ \text{func}(\{0.702381, 0.297619\})$$

- If we take a Cobb-Douglas function with parameter α, and set the error to zero, then the cause and confounder weigh in 90% as compared to their absence. PM. The left hand side with $s = 0.5$ directly evaluates to 0.5 for any α, since $0.5^{\alpha}\ (1 - 0.5)^{1-\alpha} = 0.5$.

```
eqn /. {func[x_] :> CES[1, {α, 1 - α}, x, 1, 1], Error[] → 0}
```

$$0.5 = 0.468441\ 0.488095^{\alpha}\ 0.511905^{1-\alpha} + 0.417062\ 0.297619^{1-\alpha}\ 0.702381^{\alpha}$$

```
sol = FindRoot[%, {α, 0.5}]
```

$\{\alpha \to 0.906493\}$

- If we take a linear function with parameter α, and set the error to zero, then the cause and confounder weigh in 86% as compared to their absence.

```
eqn2 = eqn /. {func[x_] :> CES[1, {α, 1 - α}, x, Infinity, 1], Error[] → 0}
```

$0.5\,(1-\alpha)+0.5\,\alpha = 0.468441\,(0.511905\,(1-\alpha)+0.488095\,\alpha)+0.417062\,(0.297619\,(1-\alpha)+0.702381\,\alpha)$

```
sol = FindRoot[%, {α, 0.5}]
```

$\{\alpha \to 0.863121\}$

The formulation in terms of marginal distributions may be inconvenient to see what this means. Let us substitute the latter α, transform back to levels by multiplication with $N = 84$, add a unit on the right hand side, and see what this means for the difference with the left hand side.

- This models $\Delta E = c_T \Delta T + c_Z \Delta Z$. This assumes that the marginals on the rhs remain the same.

```
eqn2 /. (lhs_ == rhs_) :> ((84 + 1) rhs - lhs 84)
```

$85\,(0.468441\,(0.511905\,(1-\alpha)+0.488095\,\alpha)+0.417062\,(0.297619\,(1-\alpha)+0.702381\,\alpha))-84\,(0.5\,(1-\alpha)+0.5\,\alpha)$

- This is the change in the left hand side.

```
deltaLHS = % // Simplify
```

$13.4009\,\alpha - 11.0666$

We only need to substitute the α that we found earlier. Unless we have other information on the α of course.

- Adding one unit on the right hand side means that the left hand side rises by a half unit.

```
% /. sol
```

0.5

- The particular value 0.5 may confuse. This is the proper translation to the dichotomous marginal *s*.

```
(α snew + (1 − α) (1 − snew)) (N + 1) == (α s + (1 − α) (1 − s)) N + deltaLHS
```

```
(N + 1) ((1 − snew) (1 − α) + snew α) = N (0.5 (1 − α) + 0.5 α) + 13.4009 α − 11.0666
```

```
% /. {N → 84} /. sol
```

```
85 (0.136879 (1 − snew) + 0.863121 snew) = 42.5
```

```
Solve[%, snew]
```

```
{{snew → 0.5}}
```

Or course, we have chosen α such that the lhs and rhs are equal so that the latter result should not be surprising. We might put in any regression coefficient and then still some choice of α would generate such a result. The point however is what would happen over time. A single contingency matrix provides only limited information on the dynamic relationship between the marginals. The nominal correlation coefficients and regression coefficients probably are the most on offer, but their true test comes in dynamics.

Whatever the above, the notion of "adding a unit" remains complicated for contingency tables since adding 1 to *N* percolates in all variables. In subsection 3.8 we already had a "paradigmatic example". Let us extend on it.

8. Comparison to the risk difference

8.1 A reminder

In subsection 3.8 on the preliminaries, we discussed the paradigmatic case of interpreting a regression coefficient by adding or moving a unit.

8.2 Numerical example for the 2 × 2 case

Regression coefficients can be interpreted by the effect of adding a unit (allow the total to be raised) or by moving a unit (keeping the same total). Let us focus on one regression coefficient, for example the relation between success and active treatment. Let us sum out sex to get the 2 × 2 data.

CT[Sum, "Sex", "Arthritis"]

	Active	Placebo
Some	28	14
None	13	29

We find a nominal correlation of 35% that is lower than the 39% found above, since the summed matrix neglects the effect of submatrices. In this particular numerical example, the variances appear to be about 1 so there is little difference between the correlation and regression coefficient.

NominalStatistics[mat2 = %] // N

$$\Big\{\text{ContingencyTableQ} \to \text{True},\ \text{OverallCorrelation} \to 0.357244,\ \text{Length} \to \{2.,\ 2.\},$$
$$\text{EffectiveNumberOfCategories} \to \{2.,\ 1.99887\},\ \text{Variance} \to \{1.,\ 0.999433\},$$
$$\text{BorderTotals} \to \begin{pmatrix} 42. & 42. \\ 41. & 43. \end{pmatrix},\ \text{BorderMatrices} \to \left\{\{1.,\ 2.\} \to \begin{pmatrix} 28. & 14. \\ 13. & 29. \end{pmatrix}\right\},$$
$$\text{NominalCorrelationMatrix} \to \begin{pmatrix} 1. & 0.357244 \\ 0.357244 & 1. \end{pmatrix},$$
$$\text{CovarMat} \to \begin{pmatrix} 1. & 0.357143 \\ 0.357143 & 0.999433 \end{pmatrix},\ \text{CovarRegress} \to \begin{pmatrix} 0. & 0.357345 \\ 0.357143 & 0. \end{pmatrix}\Big\}$$

Let us move one person from Active to Placebo. The special requirement is that this person is typical of the Active group when it is taken from there, but suddenly becomes typical of the Placebo group when it is put there. In other words, the effect rates per effect group are kept constant. The risk difference appears to be the same as above regression coefficient.

- The routine moves 1 from the first row to the second. But that would be the effect variable. So we first transpose.

Move1FromRow1To2[Transpose[mat2]] // N

$$\Big\{\text{Mat(In)} \to \begin{pmatrix} 28. & 13. & 41. \\ 14. & 29. & 43. \\ 42. & 42. & 84. \end{pmatrix},$$
$$\text{Mat(Out)} \to \begin{pmatrix} 27.3171 & 12.6829 & 40. \\ 14.3256 & 29.6744 & 44. \\ 41.6427 & 42.3573 & 84. \end{pmatrix},\ \text{Row}[1.] \to \{-0.682927,\ -0.317073\},$$
$$\text{Row}[2.] \to \{0.325581,\ 0.674419\},\ \text{Dif} \to \{-0.357345,\ 0.357345\}\Big\}$$

- If we do not transpose then we move a person from the "Some effect" group to the "None effect" group.

Move1FromRow1To2[mat2] // N

$$\left\{\text{Mat(In)} \to \begin{pmatrix} 28. & 14. & 42. \\ 13. & 29. & 42. \\ 41. & 43. & 84. \end{pmatrix},\right.$$

$$\text{Mat(Out)} \to \begin{pmatrix} 27.3333 & 13.6667 & 41. \\ 13.3095 & 29.6905 & 43. \\ 40.6429 & 43.3571 & 84. \end{pmatrix}, \text{Row[1.]} \to \{-0.666667, -0.333333\},$$

$$\left.\text{Row[2.]} \to \{0.309524, 0.690476\}, \text{Dif} \to \{-0.357143, 0.357143\}\right\}$$

8.3 The 2 × 2 × 2 case

8.3.1 In general

The 2 × 2 case allowed an easy formal expression. For the 2 × 2 × 2 case we now must consider the change between two variables while assuming something for the third. What to assume is not obvious by itself. The following are some assumptions that allow the recovery of the average risk differences.

Consider again the full arthritis case, simple as it is. When we shift one person from Active to Placebo, while keeping sex in the domain, then we need an assumption what happens with this variable.

CT[Show]

		Active	Placebo
Some	F	21	13
	M	7	1
None	F	6	19
	M	7	10

We can start by imposing the following conditions: (1) we order the variables, such that we move 1 person within the first variable, from row 1 to row 2, (2) the second variable keeps the same border totals, so that all impact is collected in the third variable, the effect, (3) the determinant of the border matrix of the first and second is kept the same. The latter is an approximation of the idea that this correlation would not change.

These conditions appear to be an elaborate manner to calculate the risk differences that we already determined above.

8.3.2 Putting treatment on top

CT[Order, {"Treatment", "Sex", "Effect"}]

		F	M
Active	Some	21	7
	None	6	7
Placebo	Some	13	1
	None	19	10

Move1FromRow1To2In3D[matts = % // N]

$$\{\text{Level} \to 2,\ \text{Det} \to \left\{\begin{pmatrix} 27. & 14. \\ 32. & 11. \end{pmatrix}, -151.\right\},\ \text{Mat(In)} \to \begin{pmatrix} \{21., 6.\} & \{7., 7.\} \\ \{13., 19.\} & \{1., 10.\} \end{pmatrix},$$

$$\text{Mat(Out)} \to \begin{pmatrix} \{20.4537, 5.84392\} & \{6.85119, 6.85119\} \\ \{13.2853, 19.417\} & \{1.02706, 10.2706\} \end{pmatrix},$$

$$\text{BorderMatrices(In)} \to \left\{\{1, 2\} \to \begin{pmatrix} 27. & 14. \\ 32. & 11. \end{pmatrix}, \{1, 3\} \to \begin{pmatrix} 28. & 13. \\ 14. & 29. \end{pmatrix}, \{2, 3\} \to \begin{pmatrix} 34. & 25. \\ 8. & 17. \end{pmatrix}\right\},$$

$$\text{BorderMatrices(Out)} \to \left\{\{1, 2\} \to \begin{pmatrix} 26.2976 & 13.7024 \\ 32.7024 & 11.2976 \end{pmatrix}, \{1, 3\} \to \begin{pmatrix} 27.3049 & 12.6951 \\ 14.3124 & 29.6876 \end{pmatrix},\right.$$

$$\left.\{2, 3\} \to \begin{pmatrix} 33.739 & 25.261 \\ 7.87825 & 17.1218 \end{pmatrix}\right\},\ \text{BorderTotals(In)} \to \begin{pmatrix} 41. & 43. \\ 59. & 25. \\ 42. & 42. \end{pmatrix},$$

$$\text{BorderTotals(Out)} \to \begin{pmatrix} 40. & 44. \\ 59. & 25. \\ 41.6173 & 42.3827 \end{pmatrix},\ \text{Dif} \to \begin{pmatrix} -1. & 1. \\ 0. & 3.55271 \times 10^{-15} \\ -0.382707 & 0.382707 \end{pmatrix}\}$$

- The result is exactly the risk difference R - B that we directly calculated above. The effect of 0.38 is close to the correlation coefficient of 0.39. It is not close to the regression coefficient of 0.468 calculated in the NominalStatistics. Hence, the conditions that this routine implements are for the risk difference and not for nominal correlation and nominal regression. Interestingly, though, the routine works since we impose a condition on the correlation (constant determinant).

{1, −1} . p . {f, 1 − f}

0.382707

8.3.3 Putting sex on top

CT[Order, {"Sex", "Treatment", "Effect"}]

		Active	Placebo
F	Some	21	13
	None	6	19
M	Some	7	1
	None	7	10

Move1FromRow1To2In3D[% // N]

$$\left\{\text{Level} \to 2,\ \text{Det} \to \left\{\begin{pmatrix} 27. & 32. \\ 14. & 11. \end{pmatrix}, -151.\right\},\ \text{Mat(In)} \to \begin{pmatrix} \{21., 6.\} & \{13., 19.\} \\ \{7., 7.\} & \{1., 10.\} \end{pmatrix},\right.$$

$$\text{Mat(Out)} \to \begin{pmatrix} \{20.6204, 5.89153\} & \{12.792, 18.6961\} \\ \{7.24405, 7.24405\} & \{1.04654, 10.4654\} \end{pmatrix},$$

$$\text{BorderMatrices(In)} \to \left\{\{1, 2\} \to \begin{pmatrix} 27. & 32. \\ 14. & 11. \end{pmatrix}, \{1, 3\} \to \begin{pmatrix} 34. & 25. \\ 8. & 17. \end{pmatrix}, \{2, 3\} \to \begin{pmatrix} 28. & 13. \\ 14. & 29. \end{pmatrix}\right\},$$

$$\text{BorderMatrices(Out)} \to \left\{\{1, 2\} \to \begin{pmatrix} 26.5119 & 31.4881 \\ 14.4881 & 11.5119 \end{pmatrix}, \{1, 3\} \to \begin{pmatrix} 33.4124 & 24.5876 \\ 8.29058 & 17.7094 \end{pmatrix},\right.$$

$$\left.\{2, 3\} \to \begin{pmatrix} 27.8644 & 13.1356 \\ 13.8386 & 29.1614 \end{pmatrix}\right\},\ \text{BorderTotals(In)} \to \begin{pmatrix} 59. & 25. \\ 41. & 43. \\ 42. & 42. \end{pmatrix},$$

$$\left.\text{BorderTotals(Out)} \to \begin{pmatrix} 58. & 26. \\ 41. & 43. \\ 41.703 & 42.297 \end{pmatrix},\ \text{Dif} \to \begin{pmatrix} -1. & 1. \\ 0. & 0. \\ -0.297007 & 0.297007 \end{pmatrix}\right\}$$

- The result is exactly the risk difference R_f - B_f that we directly calculated above. The effect of 0.297 is close to the correlation of 0.288. It is not close to the regression coefficient of 0.417 calculated by the routine NominalStatistics. Hence, the conditions that this routine implements are for the risk difference and not for nominal correlation and nominal regression.

{c, 1 – c} . p . {1, –1}

0.297007

8.3.4 A parallel

With logistic regression there was the case where we dropped the interaction effects and then found that this implements a restriction in the estimation that the odds ratios are the same. There was further little to be explained on this, other than that a logistic function with that parameter structure has that property. In the same way we might as well accept the various properties of the various techniques: plain vanilla regression, average risk differences seen as regression with imputed correlation, logistic regression and nominal correlation and regression.

The routine Move1FromRow1To2In3D implements some conditions in moving a unit and then recovers results from risk differences. We might continue this line of research and then recover conditions that generate the regression coefficient of nominal correlation. However, that would only mean that we do nominal correlation and regression in a more elaborate (and possibly intractable) manner. A more expedient approach is to understand how nominal correlation and regression work, what their advantages and disadvantages are. (Those are: their main focus is on variables and dimensions of higher order, and they may be useful for general ideas about association and sizes of effects.)

8.3.5 The dissimilarity between the risk difference and nominal correlation

We observed a rather small difference between the nominal correlation coefficient of 0.39 and the specific risk difference $\Delta R_{S,C} = R - B = 0.382707$. This is especially noteworthy since the risk difference would rather be a regression coefficient and not a correlation coefficient. So actually we should not explain why they aren't the same, but rather why they are so close. And we would like to see the explanation of the difference of "risk difference regression" with the nominal regression coefficient of 0.468441.

To clarify this, we need to restate the dissimilarity between the risk difference approach for categories and the nominal correlation approach for variables. The outcome of the risk difference of 0.38207 takes the category *Success* as the effect. However, there are also *variables* instead of categories, and three variables even, so that also *Treatment* and *Sex* can be considered to be the effect variable. Each choice results into a different set of average risk differences. The explanation for the arising difference between 0.382707 and 0.468441 can be seen in the point that the nominal correlation coefficient also considers these other angles. In summary, on one hand there is specific (average) risk and on the other hand there is a generalized measure of association.

We know that risk differences are summary statistics and thus liable to instability. In the ETC222 causal analysis, Colignatus (2007f), we found the true basic parameters such as *r* and *b*. The instable *R* and *B* were expressed in terms of them and the prevalence weights. Thus, it is a bit peculiar that the correlation coefficients for this particular example are close to those risk differences, and likely to be as instable. We would tend to take the correlation as more fundamental and derive the prevalence weights based upon those correlations. In itself, though, we have to think carefully about that. If a correlation has a fixed value then this might be classified as a causal phenomenon and then the ETC222 format of one cause and one confounder doesn't apply any more. Thus, if a confounder really is a confounder then we actually should not be surprised that a correlation coefficient is instable, and moves about along with the prevalence of the confounder.

9. Conclusions

In the above, we compared the suggestion of nominal correlation and nominal correlation in Colignatus (2007d) with logistic regression, benefitting from an analysis on causality in Colignatus (2007f).

Points are:

- Both approaches have their own logic. By necessity the results of nominal correlation and regression that target the variables are different from logistic regression that targets the cells.
- Nominal correlation regards the contingency table as its own correlation matrix, by proper normalization and aggregation. The approach uses the full variables, containing all categories. A correlation coefficient between two variables does not assume either of these to be the target. Correlation is done for all variables.

Regression coefficients are derived from these correlation coefficients in the same manner as for real variables. Nominal correlation doesn't use statistical independence and also corrects when there is statistical dependence. The correction is done in the same way as in OLS, by correcting for the correlation between the variables. The regression coefficients must be understood as vector regression coefficients, of which an approximate interpretation is that one moves one unit from one category to the next.

- Logistic regression takes one category as the target variable. For dichotomous variables this can be directly translated for the alternative category but this kind of translation cannot be done when there are more categories. In addition, logistic regression might neglect taking the other variables as the target, while this is standard procedure for nominal correlation. (Disciplined researchers though will run all logistic regressions.) Logistic regression does not start with statistical independence but may end with imposing it.
- The risk difference and "risk difference regression" provides a bridge to understand the difference between logistic regression and nominal regression. But the bridge only works for some distance since those three approaches different. (a) The average risk differences are not explicit about interaction effects, and apply to (independent) marginal distributions. (b) A "risk difference regression" may give about the same size of coefficients as nominal correlation, but also quite different ones when submatrices differ, and when they differ then nominal correlation explicitly handles those submatrices. (c) For the 2×2 case we considered an example that had exactly the same regression coefficient. (d) For the $2 \times 2 \times 2$ case the shift of a unit reproduces the risk difference - though different from the nominal correlation and regression coefficients. But this routine imposed the condition of equal determinants which means that it relied on the concept of nominal correlation.
- NominalCorrelation is designed for generalization (all interaction, more variables and more categories). Logistic regression is already general by itself since one can apply the format when the dimensions expand. The point thus would merely be the difference in overview and clarity.

Appendix A: Three different 2 × 2 epidemiological matrices

A.1. Three types of schemes

A.1.1 Introduction

When an economist has the objective to see whether he or she can use techniques from epidemiology for experimental economics, and when this economist then tries to understand what epidemiologists are doing, then he or she must account for the fact that there are two levels of teaching. There is one level of teaching that may not be too mathematical and that is intended e.g. for students of medicine. Those students are taught certain *molds*, recognize research formats, to plug in the data in the table, and interprete the associated statistical results. There is another level of teaching that presumes some basic mathematical abstraction. Students translate a problem into its mathematical properties, benefit from common mathematical structures, and interprete the associated statistical results but with the additional translation of mathematics back to the original problem. The economist is likely to be trained to do the latter and then he or she may not understand what the hocus pocus of the first approach is for. It can be enlightening to see that it is only a teaching format for students in epidemiology or medicine with little mathematical training or little use for mathematical abstraction.

In itself, the different *molds* are useful tools to communicate what type of problem one is discussing. That being said, those molds should not limit the use of simple mathematics, and they should not cause a distortion of true mathematical properties.

There are various epidemiological research schemes that can be cast in a 2×2 matrix. Their reasoning can be different though so that the results require a different interpretation. The general situation however is that there can be a "success" versus the absence of it. Another general point is that the relative risk and the odds ratio are both 1 when the probabilities in the matrix are independent.

Here we consider the disease-test situation, the follow-up study, and the case-control study. Our focus is on how these are taught to medical students, and we include some comments on the math.

A.1.2 The disease-test matrix.

- For a disease test: the risk is P[*diseased* | *test*] and a relative risk then is P[*diseased* | *positive*] / P[*diseased* | *negative*].

- The disease state is the cause of any outcome and hence in the columns. A random selection of the population is tested with a gold standard to find the true disease state. The same selection is submitted to a new and likely cheaper test, to find its positive and negative predictive values. Outcomes are true-positive, true-negative, false-positive and false-negative.

DiseaseTestMatrix[Table]

	Disease	Not	Tested
Positive	TP	FP	FP + TP
Negative	FN	TN	FN + TN
Sum	FN + TP	FP + TN	FN + FP + TN + TP

- This format can be compared to hypothesis testing, with the true state in the columns and the decision on acceptance or non-acceptance in the rows. You would accept the hypothesis that a person has the disease iff the test is positive.

DiseaseTestMatrix[Table, "H0 is True", "Accept H0"]

	H0 is True	¬ H0 is True	Tested
Accept H0	TP	FP	FP + TP
¬ Accept H0	FN	TN	FN + TN
Sum	FN + TP	FP + TN	FN + FP + TN + TP

- The RRisk function takes the disease-test matrix as its default, and thus with the default value of TP / (TP + FP) versus FN / (FN + TN).

RRisk[] // Simplify

$$\frac{(FN + TN)\, TP}{FN\, (FP + TP)}$$

A.1.3 The treatment-control matrix.

- For treatment or a cohort follow up study: the risk is P[*effective* | *status*] and a relative risk then is P[*effective* | *treated*] / P[*effective* | *control*].

- The cause that makes anything effective is in the columns. The researcher sets n_1 and n_2 for treatment and control. They do not have to represent population parameters.

TreatmentControlMatrix[Set, n1, n11, n2, n21];

TreatmentControlMatrix[Table]

	Effective	Ineffective	Total
Treatment	n11	n1 – n11	n1
Controls	n21	n2 – n21	n2
Sum	n11 + n21	n1 – n11 + n2 – n21	n1 + n2

- The RRiskTCM function takes the treatment-control matrix as its default. Technically it gives the same outcome as the RRisk (in terms of matrix elements divided by row sums).

RRiskTCM[] // Simplify

$$\frac{\text{n11 n2}}{\text{n1 n21}}$$

A.1.4 The case-control matrix.

- For a case-control study the focus of interest is on exposure probability P[*exposed* | *state*] and a relative measure then is P[*exposed* | *case*] / P[*exposed* | *control*]. However, since epidemiologists use the "relative risk" for the probabilities of a *disease*, they do not tend to recognize the relative exposure measure as a relative measure too.

- The cases are given as n_1 by nature (diseased who show symptoms) while n_2 are the controls (who don't show symptoms) selected by the researcher. A likely cause is identified and the exposure state is determined.

CaseControlMatrix[Table]

	Cases	Controls	Total
Exposure	n11	n12	n11 + n12
No Exposure	n1 – n11	n2 – n12	n1 – n11 – n12 + n2
Sum	n1	n2	n1 + n2

- If the case-control matrix is transposed then the relative exposure probability becomes technically the relative risk

```
RelativeExposurePr[] // Simplify
```

$$\frac{n11\, n2}{n1\, n12}$$

- But since epidemiologists think this confusing, they prefer the odds ratio.

```
OddsRatio[CaseControlMatrix[]]
```

$$\frac{n11\,(n12 - n2)}{(n11 - n1)\, n12}$$

A.1.5 Review

In sum, the scheme in this kind of teaching of epidemiology is: (a) "risk" and "relative risk" are used for the conditional probability of the disease or the cure given the state of exposure, (b) "exposure odds" and "exposure odds ratio" are used for the conditional probability (ratio) of the exposure given the state of disease. The prime advantage of this scheme is that the student would not get confused on what is conditioned on.

This approach to teaching is risky in that the emphasis in "exposure odds" switches from "exposure" to "odds". The student is trained to think "relative risk" for treatment-control and "odds" for case-control. But this is awkward since the emphasis should be on "exposure". Odds can be translated in relative probabilities. When one has an odds ratio then information on one probability allows one to directly calculate the relative probabilities. The probabilities are already given in the case-control study as well. Thus it is not obvious why one would destroy that information by switching to the odds ratio.

```
eqs = {RRisk == RelativeExposurePr[] // Simplify,
    OddsRatio == OddsRatio[CaseControlMatrix[]]}
```

$$\left\{\text{RRisk} = \frac{n11\, n2}{n1\, n12},\ \text{OddsRatio} = \frac{n11\,(n12 - n2)}{(n11 - n1)\, n12}\right\}$$

```
Solve[eqs, RRisk, n12] // Simplify
```

$$\left\{\left\{\text{RRisk} \to \frac{-\text{OddsRatio}\, n11 + n11 + n1\, \text{OddsRatio}}{n1}\right\}\right\}$$

If there is a problem in teaching the different formats of the treatment-control and case-control studies then a solution might be to use "relative exposure probabilities" to clarify what the condition is. One better changes the teaching format, since it is awkward

to hinge the distinction between relative risks and odds to the kind of condition. Epidemiologists prefer the term "risk" but they should not be afraid to use the term "probability".

It may be a bit awkward that epidemiologists put more emphasis on the disease outcome and less on the mathematical structure. However, their approach may have an advantage when one considers the issue of scaling up a small study to the level of the whole population. So there is a bit more to it than just teaching. To understand this, see the next subsection.

A.2. An example by Kleinbaum et al. (2003:106)

This example of a case-control study concerns 37 cases of diarrhea at a Haitian Resort Club suspected to have been caused by the eating of raw hamburgers. The 33 controls were a random sample from those who stayed at the same resort but who did not get diarrhea.

```
ccm = CaseControlMatrix[1, Set];
```

```
CaseControlMatrix[Table]
```

	Cases	Controls	Total
Exposure	17	7	24
No Exposure	20	26	46
Sum	37	33	70

```
OddsRatio[ccm]
```

$$\frac{221}{70}$$

Let us try to scale up the study to the larger population at risk. Let us use the exposure rates p = 17/37 and q = 7/33 respectively. With a representative sample of size N the outcome is determined by the prevalence Pr in the population, giving Pr N = 37 and implied (1 - Pr) N.

- Pr is the proportion in the population of tourists who eat bad hamburgers.

```
ccm = CaseControlMatrix[Set, Pr, Pr N, p, (1 – Pr) N, q];
```

CaseControlMatrix[Table] // Simplify

	Cases	Controls	Total
Exposure	$N\,p\,\mathrm{Pr}$	$-N\,(\mathrm{Pr}-1)\,q$	$N\,(p\,\mathrm{Pr}-q\,\mathrm{Pr}+q)$
No Exposure	$-N\,(p-1)\,\mathrm{Pr}$	$N\,(\mathrm{Pr}-1)\,(q-1)$	$-N\,(p\,\mathrm{Pr}-q\,\mathrm{Pr}+q-1)$
Sum	$N\,\mathrm{Pr}$	$N-N\,\mathrm{Pr}$	N

These still are the odds.

OddsRatio[ccm]

$$\frac{p\,(q-1)}{(p-1)\,q}$$

We now see a structural identity with the disease-test matrix. The observed 37 cases are the diseased in a proper sample and the exposure is like a test that turns up positive. Now we can use the same routine on the relative risk:

RRisk[ccm] // Simplify

$$\frac{p\,(p\,\mathrm{Pr}-q\,\mathrm{Pr}+q-1)}{(p-1)\,(p\,\mathrm{Pr}-q\,\mathrm{Pr}+q)}$$

When the prevalence drops to zero, i.e. when the event becomes extremely unlikely, the relative risk becomes the odds.

Limit[%, Pr → 0]

$$\frac{p\,(q-1)}{(p-1)\,q}$$

This latter can be explained as follows. With proper sample sizes, the RRisk gets its proper value, different from the OddsRatio, since RRisk also considers the probability of being exposed. When we don't have any information on the prevalence then the odds ratio still gives useful information - and it even gives information when the prevalence would be zero.

In sum, epidemiologists prefer to work with tables with above structure, such that when a case-control study is scaled up to the population, then the standard definition of relative risk, using the rows, becomes meaningful. When the type of data do not fit the right table then they regard it as confusing to discuss a relative exposure probability and henceforth they prefer the odds ratio.

This is what they do. But they don't explain it in that manner - leaving one to wonder why they don't use the same mathematics on relative rates ...

Because, mathematically, when we transpose the case-control matrix then we get an expression for a relative probability, that has the same mathematical properties as the relative risk (but not the same epidemiological interpretation on what it conditioned on).

```
CaseControlMatrix[] // Transpose // RRisk // Simplify
```

$$\frac{p}{q}$$

Concluding, we can observe that there is a certain system in these schemes in epidemiology. The kinds of teaching, the need for easy mnemonics, the low information requirement of the odds ratio, the option to scale up results to the population level. Yet, the mathematical approach has all these advantages too, plus more. So it may be a key question why one would think that students in medicine should not be able to master a little bit of mathematics too.

A.3. Linking up logistic regression and the odds ratio

As said, some epidemiologists have been trained to use the exposure odds ratio for the case-control study. To link up logistic regression to the odds ratio, we may observe that logistic regression uses odds and log-odds, so that odds ratio's are only calculated afterwards. From mere mathematical transformation it is immaterial whether we use probabilities or odds, and this would also hold for relative probabilities and odds ratios (assuming that we know the marginals). But it might matter a great deal which of the latter is taken as constant.

- Relations between probabilities, relative risk, odds, odds ratio.

```
eqs = {RRisk == Pr[1] / Pr[2], Odds1 == Odds[Pr[1]],
    Odds2 == Odds[Pr[2]], OddsRatio == Odds1 / Odds2}
```

$$\left\{\text{RRisk} = \frac{\text{Pr}(1)}{\text{Pr}(2)}, \text{Odds1} = \frac{\text{Pr}(1)}{1-\text{Pr}(1)}, \text{Odds2} = \frac{\text{Pr}(2)}{1-\text{Pr}(2)}, \text{OddsRatio} = \frac{\text{Odds1}}{\text{Odds2}}\right\}$$

```
Solve[eqs, {RRisk, OddsRatio}, {Pr[1], Pr[2]}] // Simplify
```

$$\left\{\left\{\text{OddsRatio} \to \frac{\text{Odds1}}{\text{Odds2}}, \text{RRisk} \to \frac{\text{Odds2 Odds1} + \text{Odds1}}{\text{Odds1 Odds2} + \text{Odds2}}\right\}\right\}$$

The point thus is that if Pr[1] is free to move about, then we either fix RRisk or OddsRatio, unless they are both 1. (For the case-control situation we can translate this to the ratio of exposure probabilities.)

```
Solve[eqs, {RRisk}, {Pr[2], Odds1, Odds2}] // Simplify
```

$$\{\{\text{RRisk} \to -\text{Pr}(1)\,\text{OddsRatio} + \text{OddsRatio} + \text{Pr}(1)\}\}$$

There is also the risk difference. The above can be repeated for that.

```
eqs = {RiskDiff == Pr[1] - Pr[2], Odds1 == Odds[Pr[1]],
    Odds2 == Odds[Pr[2]], OddsRatio == Odds1 / Odds2}
```

$$\left\{\text{RiskDiff} = \text{Pr}(1) - \text{Pr}(2),\ \text{Odds1} = \frac{\text{Pr}(1)}{1 - \text{Pr}(1)},\ \text{Odds2} = \frac{\text{Pr}(2)}{1 - \text{Pr}(2)},\ \text{OddsRatio} = \frac{\text{Odds1}}{\text{Odds2}}\right\}$$

```
Solve[eqs, {RiskDiff, OddsRatio}, {Pr[1], Pr[2]}] // Simplify
```

$$\left\{\left\{\text{RiskDiff} \to \frac{\text{Odds1} - \text{Odds2}}{\text{Odds2}\,\text{Odds1} + \text{Odds1} + \text{Odds2} + 1},\ \text{OddsRatio} \to \frac{\text{Odds1}}{\text{Odds2}}\right\}\right\}$$

```
Solve[eqs, {RiskDiff}, {Pr[2], Odds1, Odds2}] // Simplify
```

$$\left\{\left\{\text{RiskDiff} \to \frac{(\text{OddsRatio} - 1)\,(\text{Pr}(1) - 1)\,\text{Pr}(1)}{\text{OddsRatio}\,(\text{Pr}(1) - 1) - \text{Pr}(1)}\right\}\right\}$$

It is just as simple to translate the log-odds from a logistic regression to the probabilities and directly substitute them in the risk difference. This is also a cleaner method, since it avoids a possible confusion on the role of the exposure odds ratio. The key point remains that we should always be aware what probabilities we are discussing, and what they are conditioned on.

Above tables are all 2 × 2. Logistic regression may be a bit overdone there. The method comes into the picture when there would be a third variable, e.g. a confounder.

A.4. Relation to causality

The discussion on causality in the 2 × 2 × 2 case by Colignatus (2007f) assumes that the marginal distributions of the cause and the confounder can be meaningfully interpreted, either for observational studies outside of control by the researcher or by controlled trial. The above clarifies that there are different contexts of interpretation, i.e. disease-test, (follow-up) treatment-control, or case-control. In such studies there may be a confluence of two causes, e.g. a disease and a test that jointly produce an effect, while Colignatus (2007) only studies an effect, a cause and a confounder.

Appendix B: Aggregation, using the example of the CES function

Economics has a strong tradition in aggregation. Index numbers, the functions for utility, consumption, production, and social welfare, and also issues in voting.

An example is given by the "constant elasticity of substitution" (CES) aggregator function. The function has nonnegative parameters: the level A, factor weights c_i per variable x_i, a possible "returns to scale" v, and the elasticity of substitution σ. The aggregator function takes particular shapes for particular values $\sigma = 0, 1, \infty$.

```
Clear[A, c, x, v]
```

- The Leontief function ($\sigma = 0$, $v = 1$)

```
y == CES[A, {c1, c2}, {x1, x2}, 0]
```

$$y = A \operatorname{Min}\left(\frac{x_1}{c_1}, \frac{x_2}{c_2}\right)$$

- The Cobb-Douglas function ($\sigma = 1$, $v = 1$)

```
y == CES[A, {c1, c2}, {x1, x2}, 1]
```

$$y = A\, x_1^{c_1}\, x_2^{c_2}$$

- Line: infinite substitutionability ($\sigma = \infty$, $v = 1$)

```
y == CES[A, {c1, c2}, {x1, x2}, ∞]
```

$$y = A\,(c_1\, x_1 + c_2\, x_2)$$

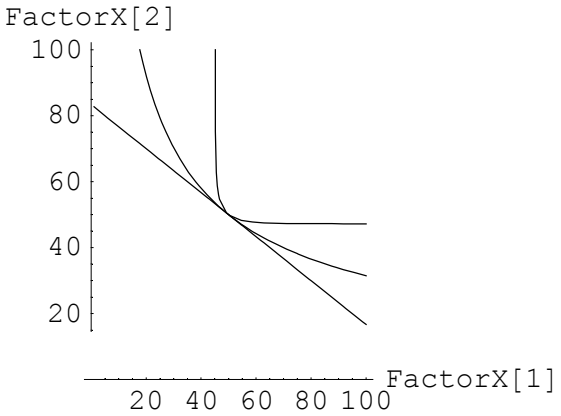
- The full CES function ($\sigma = S$, also including scale parameter v)

```
y == CES[A, {c1, c2}, {x1, x2}, S, v]
```

$$y = A\left(c_1\, x_1^{1-\frac{1}{S}} + c_2\, x_2^{1-\frac{1}{S}}\right)^{\frac{v}{1-\frac{1}{S}}}$$

- Assuming a constant level of y then the contours of x_1 and x_2 clarify the impact of values of σ. The contour plot also clarifies σ's name of "elasticity of substitution": how much one would have to sacrifice of one factor to gain the other factor (while remaining on the contour).

```
SetOptions[CES, Constant → 1, FactorCoefficients → {.4, .6}];
CESContours[{50}, 100, {Ces → 0.1}, {Ces → 1},
    {Ces → ∞}, AxesOrigin → {0, 0}, TextStyle → {FontSize → 11}];
```



One may recognize the Box-Cox transformation: $f[x, \lambda] = (x^{\lambda} - 1) / \lambda$ for $\lambda \neq 0$ and $f[x, 0] = \text{Log}[x]$. Thus output $f[y, \lambda] = \Sigma\, c_i\, f[x_i, \lambda]$.

The economics literature is abound with criticism on aggregation. Macro outcomes are only accurate when the micro developments are proportional, and since they seldom are, and since many researchers wish to reduce the error, there is an ever increasing emphasis on micro modelling. However, there still remains a need to keep an overview and approximate macro functions still serve good purposes.

For the current case at hand, nominal correlation, it is important to note that an aggregator here uses nominal categories, that have no numerical values, so that aggregation concerns the frequencies and not an aggregation-weighed average of scores with the frequencies as weights. For nominal regression it might suffice that one uses only the variances of the frequencies, as discussed in Colignatus (2007d).

Appendix C: Examples from Garson (2007b)

C.1 Introduction

Garson (2007b) contains two clarifying examples on the analysis of contingency tables. We can provide the nominal correlation matrices and implied regression coefficients to see what their added value is. The first case on literacy is illuminating on the kind of causes behind a summary table, the second case will be one with little correlation but (still) statistically significant parameters.

C.2 The literacy issue

C.2.1 Introduction

This example starts with this summary table. Garson (2007b) puts non-literacy on top but we will regard non-literacy as the default background situation that causally requires additional effort to turn into literacy. The table actually is between the effect and the confounder. The table suggests a connection between between race and literacy, which must be due to a cause not shown, such as regional differences, and which the nazi type of researcher would ascribe to genetics.

```
TableForm[{{2, 6}, {6, 2}},
  TableHeadings → {{"Literate", Not["Literate"]}, {"Black", "White"}}]
```

	Black	White
Literate	2	6
¬ Literate	6	2

Garson (2007b): “In the traditional elaboration model using crosstabulation, one divided an original table (ex., literacy by race) into two or more subtables based on the control variable (ex., subtables for region = South or North). For instance, the overall table might show an association between race=black and literacy=not, but if the South tended to have more blacks than the North and more non-literates for both races, then the original relation might well be spurious because region was a control variable. In general, the control variable subtables migh[t] show the same relationship as the original table, no relationship, or a relationship for one subtable but not the other.”

Indeed, we will see three full tables (Model A, B and C) with *Effect* = literacy, *Truth* = region and *Confounding* = race, that all have the same above summary result. These tables illuminate that entirely different processes can generate the same result so that one should be careful in drawing conclusion without knowing all data.

In Garson's use of language, the region is called the "control variable". For us, the region will contain the cause, or at least the proxy of the cause, such as the different mixes of populations, quality of education, and cultural habits such as discrimination and work ethics. Since we maintain the ETC order, the data below will be ordered differently than by Garson.

Garson (2007b): "The original table above is shown with three different possible splits by the control variable Region. In Model A, the split tables have the same relationship as the original table. There is no control effect (...). In Model B, there is full explanation (total control by the control variable Region) and each component of the loglinear generating class is an interaction involving the control variable). In Model C, the original relationship disappears (is controlled) in the South region but is stronger than the original in the North region, showing the original table to be a misleading average. For Model C, the loglinear generating class contains all three interactions."

PM. One approach would be to take these data as sacrosanct and derive strong conclusions on those. Given that they are made up it is better to treat them as such. Garson (2007) provides the context where they would be used for real conclusions and we have the role to debunk that.

C.2.2 Model A

Since we reorder the data it may be more difficult to see that they have the "same structure" as the summary table. In our presentation, it is immediately clear that the regions have the same distribution. It would still be the case that all differences in literacy can only be explained by some hidden cause, but people with a prejudice might employ the logical fallacy "since there are no regionial differences it must be genetitcs".

CT[Set, "Garson: Model A"]

		South	North
Literate	Black	1	1
	White	3	3
¬ Literate	Black	3	3
	White	1	1

- Summing out the region gives the original summary table.

```
CT[Sum, "Region", "Garson: Model A"]
```

	Black	White
Literate	2	6
¬ Literate	6	2

- Reordering generates the "same structure" that Garson (2007b) referred to.

```
CT[Order, {"Region", "Race", "Literacy"}, "Garson: Model A"]
```

		Black	White
South	Literate	1	3
	¬ Literate	3	1
North	Literate	1	3
	¬ Literate	3	1

- It is instructive to see the correlations between race and literacy in the separate regions.

```
NominalCorrelation /@ %
```

$$\left\{-\frac{1}{2}, -\frac{1}{2}\right\}$$

Nominal correlation shows that region is not correlated with literacy and race.

```
TableForm[
  NominalCorrelationMatrix[CT[ "Garson: Model A", Data]], TableHeadings →
    {CT[ "Garson: Model A", Variables], CT[ "Garson: Model A", Variables]}]
```

	Literacy	Region	Race
Literacy	1	0	$-\frac{1}{2}$
Region	0	1	0
Race	$-\frac{1}{2}$	0	1

NominalStatistics[CT["Garson: Model A", Data]]

$$\Big\{\text{ContingencyTableQ} \to \text{True},\ \text{OverallCorrelation} \to 0.5,\ \text{Length} \to \{2, 2, 2\},$$

$$\text{EffectiveNumberOfCategories} \to \{2., 2., 2.\},\ \text{Variance} \to \{1., 1., 1.\},\ \text{BorderTotals} \to \begin{pmatrix} 8 & 8 \\ 8 & 8 \\ 8 & 8 \end{pmatrix},$$

$$\text{BorderMatrices} \to \left\{\{1, 2\} \to \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix},\ \{1, 3\} \to \begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix},\ \{2, 3\} \to \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}\right\},$$

$$\text{NominalCorrelationMatrix} \to \begin{pmatrix} 1. & 0. & -0.5 \\ 0. & 1. & 0. \\ -0.5 & 0. & 1. \end{pmatrix},$$

$$\text{CovarMat} \to \begin{pmatrix} 1. & 0. & -0.5 \\ 0. & 1. & 0. \\ -0.5 & 0. & 1. \end{pmatrix},\ \text{CovarRegress} \to \begin{pmatrix} 0. & 0. & -0.5 \\ 0. & 0. & 0. \\ -0.5 & 0. & 0. \end{pmatrix}\Big\}$$

The causal analysis (notably looking at *r, b, w, v*) shows that the confouder has a counter-intuitive impact, i.e. *r* and *b* are 0.75 but when the confounder is present they drop to *w* and *v* of 0.25. There is some hidden cause not shown, e.g. data selection.

(ETCStatistics[CT["Garson: Model A", Data]] // N) // MatrixForm

Matrix ETCStatistics["Cause, True, Ratio"]

	Cause	¬ Cause	Total
Success	0.375	0.375	0.75
¬ Success	0.125	0.125	0.25
Sum	0.5	0.5	1.

Matrix ETCStatistics["Cause"]

	Cause	¬ Cause	Total
Success	4	4	8
¬ Success	4	4	8
Sum	8	8	16

Matrix ETCStatistics["Confounder"]

	Cause	¬ Cause	Total
Confounder	4	4	8
¬ Confounder	4	4	8
Sum	8	8	16

Matrix ETCStatistics["Seeming"]

	Confounder	¬ Confounder	Total
Success	2	6	8
¬ Success	6	2	8
Sum	8	8	16

$$\left(\begin{array}{l}
N \to 16. \\
\text{NSuccess} \to 8. \\
\text{NCause} \to 8. \\
\text{NConfounder} \to 8. \\
\text{MarginalPr(Success)} \to 0.5 \\
\text{MarginalPr(Cause)} \to 0.5 \\
\text{MarginalPr(Confounder)} \to 0.5 \\
\text{IndependentPr(Truth, Confounding)} \to \text{True} \\
(\text{Success} \perp \neg\ \text{Confounder})(\text{Cause}) \to \text{False} \\
(\text{Success} \perp \neg\ \text{Confounder})(\neg\ \text{Cause}) \to \text{False} \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \text{Cause}, \neg\ \text{Confounder}] \to 0.75 \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \neg\ \text{Cause}, \neg\ \text{Confounder}] \to 0.75 \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \text{Cause}, \text{Confounder}] \to 0.25 \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \neg\ \text{Cause}, \text{Confounder}] \to 0.25 \\
\text{Risk} \to \begin{pmatrix} 0.25 & 0.25 \\ 0.75 & 0.75 \end{pmatrix} \\
\text{Interaction} \to \{\text{Add} \to 0., \text{Times} \to 0.\} \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \text{Cause}] \to 0.5 \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \neg\ \text{Cause}] \to 0.5 \\
\text{ConditionalPr}[\ \text{Cause}\ ][\ \text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\ \text{Cause}\ ][\ \neg\ \text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \text{Confounder}] \to 0.25 \\
\text{ConditionalPr}[\ \text{Success}\ ][\ \neg\ \text{Confounder}] \to 0.75 \\
\text{RRisk(True)} \to 1. \\
\text{RRisk(Cause)} \to 1. \\
\text{RelativePr(Confounder)} \to 1. \\
\text{RRisk(Seeming)} \to 0.333333 \\
\text{ETCAdjustedRRisk} \to \{1., 1., 1., 1.\} \\
\text{Conditions} \to \{\text{False}, \text{False}, \text{True}, \text{False}, \text{True}\} \\
\text{ConditionalPr}[\ \neg\ \text{Success}\ ][\ \text{Cause}, \neg\ \text{Confounder}] \to 0.25 \\
\text{ConditionalPr}[\ \neg\ \text{Success}\ ][\ \neg\ \text{Cause}, \neg\ \text{Confounder}] \to 0.25 \\
\text{ConditionalPr}[\ \neg\ \text{Success}\ ][\ \text{Cause}, \text{Confounder}] \to 0.75 \\
\text{ConditionalPr}[\ \neg\ \text{Success}\ ][\ \neg\ \text{Cause}, \text{Confounder}] \to 0.75 \\
\text{Safety} \to \begin{pmatrix} 0.75 & 0.75 \\ 0.25 & 0.25 \end{pmatrix} \\
\text{SimpleCauseQ} \to \begin{pmatrix} \text{False} & \text{False} \\ \text{False} & \text{False} \end{pmatrix} \\
\text{ETCSimpson} \to \{\text{Necessary} \to \text{False}, \text{Sufficient} \to \{\text{False}, \text{False}, \text{False}\}\}
\end{array}\right)$$

C.2.3 Model B

Recall the diagnosis by Garson (2007): "In Model B, there is full explanation (total control by the control variable Region) and each component of the loglinear generating class is an interaction involving the control variable)."

```
CT[Set, "Garson: Model B"]
```

		South	North
Literate	Black	0	2
	White	0	6
¬ Literate	Black	6	0
	White	2	0

- Summing out the region gives the original summary table.

```
CT[Sum, "Region", "Garson: Model B"]
```

	Black	White
Literate	2	6
¬ Literate	6	2

- Reordering in the Garson (2007b) format.

```
CT[Order, {"Region", "Race", "Literacy"}, "Garson: Model B"]
```

		Black	White
South	Literate	0	0
	¬ Literate	6	2
North	Literate	2	6
	¬ Literate	0	0

The model is a bit peculiar since the South would have no literates and the North would have only literates. Nominal correlation shows that this is painful.

```
TableForm[
  NominalCorrelationMatrix[CT[ "Garson: Model B", Data]], TableHeadings →
    {CT[ "Garson: Model B", Variables], CT[ "Garson: Model B", Variables]}]
```

	Literacy	Region	Race
Literacy	1	−1	Indeterminate
Region	−1	1	Indeterminate
Race	Indeterminate	Indeterminate	1

There is no need to try the other routines. In causal analysis we would need to explain where this extreme difference in distribution comes from.

C.2.4 Model C

Recall the diagnosis by Garson (2007): "In Model C, the original relationship disappears (is controlled) in the South region but is stronger than the original in the North region, showing the original table to be a misleading average. For Model C, the loglinear generating class contains all three interactions."

- The columns for the South and North differ. Since the South is uniform, all effects of the summary table must be generated by the North. The peculiarity of Model B is retained for the North only.

  CT[Set, "Garson: Model C"]

		South	North
Literate	Black	2	0
	White	2	4
¬ Literate	Black	2	4
	White	2	0

- Summing out the region gives the original summary table.

  CT[Sum, "Region", "Garson: Model C"]

	Black	White
Literate	2	6
¬ Literate	6	2

- Reordering in the Garson (2007b) format.

  CT[Order, {"Region", "Race", "Literacy"}, "Garson: Model C"]

		Black	White
South	Literate	2	2
	¬ Literate	2	2
North	Literate	0	4
	¬ Literate	4	0

- It is instructive to see the correlations between literacy and race in the separate regions.

  NominalCorrelation /@ %

  {0, −1}

Nominal correlation shows that region is not correlated with literacy and race.

```
TableForm[
  NominalCorrelationMatrix[CT[ "Garson: Model C", Data]], TableHeadings →
    {CT[ "Garson: Model C", Variables], CT[ "Garson: Model C", Variables]}]
```

	Literacy	Region	Race
Literacy	1	0	$-\frac{1}{2}$
Region	0	1	0
Race	$-\frac{1}{2}$	0	1

For nominal statistics, there is no real difference between Model A and Model C. There seemed to be a difference but this is exposed by considering the correlations of the submatrices. The pecularity of the North still makes this a nazi scenario.

```
NominalStatistics[CT[ "Garson: Model C", Data]]
```

$$\left\{\text{ContingencyTableQ} \to \text{True}, \text{OverallCorrelation} \to 0.5, \text{Length} \to \{2, 2, 2\},\right.$$

$$\text{EffectiveNumberOfCategories} \to \{2., 2., 2.\}, \text{Variance} \to \{1., 1., 1.\}, \text{BorderTotals} \to \begin{pmatrix} 8 & 8 \\ 8 & 8 \\ 8 & 8 \end{pmatrix},$$

$$\text{BorderMatrices} \to \left\{\{1, 2\} \to \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}, \{1, 3\} \to \begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix}, \{2, 3\} \to \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}\right\},$$

$$\text{NominalCorrelationMatrix} \to \begin{pmatrix} 1. & 0. & -0.5 \\ 0. & 1. & 0. \\ -0.5 & 0. & 1. \end{pmatrix},$$

$$\left.\text{CovarMat} \to \begin{pmatrix} 1. & 0. & -0.5 \\ 0. & 1. & 0. \\ -0.5 & 0. & 1. \end{pmatrix}, \text{CovarRegress} \to \begin{pmatrix} 0. & 0. & -0.5 \\ 0. & 0. & 0. \\ -0.5 & 0. & 0. \end{pmatrix}\right\}$$

The causal analysis confirms that in the North all whites benefit from education while none of the blacks do ($b = 1$, $v = 0$). There must be a hidden explanation that is not shown in these tables.

```
(ETCStatistics[CT[ "Garson: Model C", Data], Print → False] // N) // MatrixForm
```

$$\begin{pmatrix}
N \to 16. \\
\text{NSuccess} \to 8. \\
\text{NCause} \to 8. \\
\text{NConfounder} \to 8. \\
\text{MarginalPr(Success)} \to 0.5 \\
\text{MarginalPr(Cause)} \to 0.5 \\
\text{MarginalPr(Confounder)} \to 0.5 \\
\text{IndependentPr(Truth, Confounding)} \to \text{True} \\
(\text{Success} \perp \neg\,\text{Confounder})(\text{Cause}) \to \text{True} \\
(\text{Success} \perp \neg\,\text{Confounder})(\neg\,\text{Cause}) \to \text{False} \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Cause}, \neg\,\text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Cause}, \neg\,\text{Confounder}] \to 1. \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Cause}, \text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Cause}, \text{Confounder}] \to 0. \\
\text{Risk} \to \begin{pmatrix} 0.5 & 0. \\ 0.5 & 1. \end{pmatrix} \\
\text{Interaction} \to \{\text{Add} \to 1., \text{Times} \to \text{Indeterminate}\} \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Cause}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Cause}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Cause}\,][\,\text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Cause}\,][\,\neg\,\text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\text{Confounder}] \to 0.25 \\
\text{ConditionalPr}[\,\text{Success}\,][\,\neg\,\text{Confounder}] \to 0.75 \\
\text{RRisk(True)} \to 0.5 \\
\text{RRisk(Cause)} \to 1. \\
\text{RelativePr(Confounder)} \to 1. \\
\text{RRisk(Seeming)} \to 0.333333 \\
\text{ETCAdjustedRRisk} \to \{1., 0.5, \infty, \infty\} \\
\text{Conditions} \to \{\text{False}, \text{False}, \text{True}, \text{False}, \text{True}\} \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\text{Cause}, \neg\,\text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\neg\,\text{Cause}, \neg\,\text{Confounder}] \to 0. \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\text{Cause}, \text{Confounder}] \to 0.5 \\
\text{ConditionalPr}[\,\neg\,\text{Success}\,][\,\neg\,\text{Cause}, \text{Confounder}] \to 1. \\
\text{Safety} \to \begin{pmatrix} 0.5 & 1. \\ 0.5 & 0. \end{pmatrix} \\
\text{SimpleCauseQ} \to \begin{pmatrix} \text{False} & \text{True} \\ \text{False} & \text{False} \end{pmatrix} \\
\text{ETCSimpson} \to \{\text{Necessary} \to \text{False}, \text{Sufficient} \to \{\text{True}, \text{False}, \text{False}\}\}
\end{pmatrix}$$

C.2.5 Conclusion

(1) It is useful to see that a marginal table can have different kinds of submatrices. (2) Nominal correlation and causal analysis help in the analysis. (3) It may be that researchers versed in logistic regression arrive at the same kind of conclusion, only phrased differently. In that case, nominal correlation and causal analysis can help them in communication with researchers who are used to real-valued data.

C.3 The party, race and gender issue

This is the table.

CT[Set, "Garson: Party, Race, Gender"]

		White	Hispanic	Black
Male	Democrat	40	56	87
	Independent	21	23	14
	Republican	62	41	38
Female	Democrat	51	66	98
	Independent	24	18	11
	Republican	39	27	21

The correlations are small. This is an example where logistic regression can generate statistically significant results that would not need to be so significant when judged by content.

```
NominalStatistics[CT["Garson: Party, Race, Gender", Data]] // N // Chop
```

$$\Big\{\text{ContingencyTableQ} \to \text{True},\ \text{OverallCorrelation} \to 0.151853,$$
$$\text{Length} \to \{2., 3., 3.\},\ \text{EffectiveNumberOfCategories} \to \{1.99732, 2.98623, 2.43892\},$$
$$\text{Variance} \to \{0.998658, 0.993285, 0.679363\},$$
$$\text{BorderTotals} \to \{\{382., 355.\}, \{237., 231., 269.\}, \{398., 111., 228.\}\},\ \text{BorderMatrices} \to$$
$$\left\{\{1., 2.\} \to \begin{pmatrix} 123. & 120. & 139. \\ 114. & 111. & 130. \end{pmatrix},\ \{1., 3.\} \to \begin{pmatrix} 183. & 58. & 141. \\ 215. & 53. & 87. \end{pmatrix},\ \{2., 3.\} \to \begin{pmatrix} 91. & 45. & 101. \\ 122. & 41. & 68. \\ 185. & 25. & 59. \end{pmatrix}\right\},$$
$$\text{NominalCorrelationMatrix} \to \begin{pmatrix} 1. & 0.0370336 & 0.146808 \\ 0.0370336 & 1. & -0.0121359 \\ 0.146808 & -0.0121359 & 1. \end{pmatrix},$$
$$\text{CovarMat} \to \begin{pmatrix} 0.998658 & 0.0368843 & 0.120923 \\ 0.0368843 & 0.993285 & -0.00996917 \\ 0.120923 & -0.00996917 & 0.679363 \end{pmatrix},$$
$$\text{CovarRegress} \to \begin{pmatrix} 0 & 0.0389258 & 0.178565 \\ 0.0395634 & 0 & -0.0217163 \\ 0.121623 & -0.0145529 & 0 \end{pmatrix}\Big\}$$

Appendix D: Plain vanilla regression for the 2 × 2 × 2 case

Plain vanilla regression may generate similar coefficients as the equation $S = \{C, 1 - C\}$. p . $\{F, 1 - F\}$. In the general case there will be an error, that also might cause different coefficients. However, under marginal independence of the cause and the confouder, the error will be zero, also generating the same coefficients.

Consider the general 2 × 2 × 2 case.

```
Clear[a, b, c, d, e, f, g, h]
```

```
lis = {{{a, b}, {c, d}}, {{e, f}, {g, h}}}
```

$$\begin{pmatrix} \{a, b\} & \{c, d\} \\ \{e, f\} & \{g, h\} \end{pmatrix}$$

```
TableForm[lis]
```

$$\begin{matrix} a & c \\ b & d \end{matrix}$$

$$\begin{matrix} e & g \\ f & h \end{matrix}$$

```
p = lis[[1]] / (lis[[1]] + lis[[2]])
```

$$\begin{pmatrix} \frac{a}{a+e} & \frac{b}{b+f} \\ \frac{c}{c+g} & \frac{d}{d+h} \end{pmatrix}$$

The marginals are:

```
{sm, cm, fm} = (First /@ BorderTotals[lis]) / Add[lis] // N
```

$$\left\{\frac{a+b+c+d}{a+b+c+d+e+f+g+h}, \frac{a+b+e+f}{a+b+c+d+e+f+g+h}, \frac{a+c+e+g}{a+b+c+d+e+f+g+h}\right\}$$

The error in estimating the marginal on S from the probabilities and assumed independent marginals for C and F:

```
Error[] == S - {C, 1 - C} . p . {F, 1 - F}
```

$$\text{Error}() = -F\left(\frac{c\,(1-C)}{c+g} + \frac{a\,C}{a+e}\right) - (1-F)\left(\frac{b\,C}{b+f} + \frac{(1-C)\,d}{d+h}\right) + S$$

These will be the coefficients.

```
term = Collect[{C, 1 - C} . p . {F, 1 - F}, {C, F, C F}]
```

$$\frac{d}{d+h} + F\left(\frac{c}{c+g} - \frac{d}{d+h}\right) + C\left(\frac{b}{b+f} + F\left(\frac{a}{a+e} - \frac{b}{b+f} - \frac{c}{c+g} + \frac{d}{d+h}\right) - \frac{d}{d+h}\right)$$

```
Coefficient[term, F C]
```

$$\frac{a}{a+e} - \frac{b}{b+f} - \frac{c}{c+g} + \frac{d}{d+h}$$

```
Coefficient[term, C] - %F // Simplify
```

$$\frac{b}{b+f} - \frac{d}{d+h}$$

```
Coefficient[term, F] - %%C // Simplify
```

$$\frac{c\,h - d\,g}{(c+g)\,(d+h)}$$

This allows us to determine the error.

```
Error[] → S – Collect[{C, 1 – C}.p.{F, 1 – F}, {C, F, C F}]
```

Error() →

$$-\frac{d}{d+h}-F\left(\frac{c}{c+g}-\frac{d}{d+h}\right)-C\left(\frac{b}{b+f}+F\left(\frac{a}{a+e}-\frac{b}{b+f}-\frac{c}{c+g}+\frac{d}{d+h}\right)-\frac{d}{d+h}\right)+S$$

```
sol = % /. {S → sm, F → fm, C → cm};
```

```
sol2 = FullSimplify[sol, Assumptions → Thread[{a, b, c, d, e, f, g, h} ≥ 0]]
```

$$\text{Error}() \to (((b+f)(c+g)-(a+e)(d+h))$$
$$(b(-a\,d\,g+e\,h\,g+c(d\,e+2\,h\,e+a\,h))-f(a\,c\,d+2\,a\,g\,d+e\,g\,d-c\,e\,h+a\,g\,h)))/$$
$$((a+e)(b+f)(c+g)(d+h)(a+b+c+d+e+f+g+h)^2)$$

And the error is null when the numerator is null too.

```
part1 = First[Numerator[Error[] /. sol2]]
```

$$(b+f)(c+g)-(a+e)(d+h)$$

There are two cases:

```
part1 == 0
```

$$(b+f)(c+g)-(a+e)(d+h)=0$$

```
part2 = FullSimplify[ Numerator[ Error[] /. sol2] / part1 == 0,
    Assumptions → Thread[{a, b, c, d, e, f, g, h} ≥ 0]]
```

$$f(a\,c\,d+2\,a\,g\,d+e\,g\,d-c\,e\,h+a\,g\,h)=b(-a\,d\,g+e\,h\,g+c(d\,e+2\,h\,e+a\,h))$$

The first case is true when the base of the probabilities is independent - i.e. the border matrix that arises by summing out the success, which gives the marginal for both cause and confounder, would have a determinant of zero. We would summarize this by saying that cause and confounder are distributed independently - but we should be careful to note that this only holds for their joint marginal and not for the inner matrices.

```
lis[[1]] + lis[[2]]
```

$$\begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix}$$

```
Det[%] == 0 // Simplify
```

$$(a+e)(d+h)=(b+f)(c+g)$$

The second case is not further looked into.

Expand /@ part2

$$acdf + 2adgf + degf - cehf + aghf = bcde + 2bche + bghe - abdg + abch$$

Literature

Colignatus is the name of Thomas Cool in science.

Christensen, R. (1997), "Log-Linear Models and Logistic Regression", Springer, http://www.math.unm.edu/~fletcher/llm.html, see http://books.google.nl/books?id=7acdFD_eX24C&dq=Log-Linear+Models+and+Logistic+Regression etcetera (only partly available on the web, non-retrievable)

Cool, Th. (1999, 2001), "The Economics Pack, Applications for *Mathematica*", http://www.dataweb.nl/~cool, ISBN 90-804774-1-9, JEL-99-0820

Colignatus, Th. (2006), "On the sample distribution of the adjusted coefficient of determination (R2Adj) in OLS", http://library.wolfram.com/infocenter/MathSource/6269/

Colignatus, Th. (2007a), "A logic of exceptions", http://www.dataweb.nl/~cool, ISBN 978-90-804774-4-5

Colignatus, Th. (2007b), "Voting theory for democracy", 2nd edition, http://www.dataweb.nl/~cool, ISBN 978-90-804774-5-2

Colignatus, Th. (2007c), "A measure of association (correlation) in nominal data (contingency tables), using determinants", a earlier version (3rd publishable draft), http://ideas.repec.org/p/pra/mprapa/2662.html

Colignatus, Th. (2007d), "Correlation in contingency tables. A measure of association or correlation in nominal data (contingency tables), using determinants", the improved version of Colignatus (2007c), but useful to mention in this list of references if only an abridged version is eventually published, http://mpra.ub.uni-muenchen.de/3394/

Colignatus, Th. (2007e), "Elementary statistics and causality", work in progress, http://www.dataweb.nl/~cool/Papers/ESAC/Index.html

Colignatus, Th. (2007f), "The 2 × 2 × 2 case in causality, of an effect, a cause and a confounder", http://mpra.ub.uni-muenchen.de/3351/, Retrieved from source

Friendly, M. (2007), "Categorical Data Analysis with Graphics", Retrieved from http://www.math.yorku.ca/SCS/Courses/grcat/grc6.html (citing the data from Koch & Stokes (1991))

Garson, D. (2007a), "Logistic Regression", http://www2.chass.ncsu.edu/garson/pa765/logistic.htm, Retrieved from source

Garson, D. (2007b), "Log-Linear, Logit, and Probit Models", http://www2.chass.ncsu.edu/garson/pa765/logit.htm, Retrieved from source

Kleinbaum, D.G., K.M. Sullivan and N.D. Barker (2003), "ActivEpi Companion texbook", Springer

Lowry, R. (2007), "VassarStats. Simple logistic regression", website, http://faculty.vassar.edu/lowry/logreg1.html, Retrieved from Source

Social Research Methods (2007), "The 2*2*2 Contingency Table", http://www.socialresearchmethods.net/tutorial/Cho/222table.htm, Retrieved from source

Theil H. (1971), "Principles of econometrics", North-Holland

Weisstein, Eric W. (2007) "Fisher's Exact Test." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/FishersExactTest.html

Noted but not consumed yet

PM. The author is an econometrician and may be assumed to know various techniques in the below. These publications would still need to be digested to determine what is known and what not.

Agresti, A. (2002), "Categorical Data Analysis", 2nd Edition, Wiley, http://www.stat.ufl.edu/~aa/

Agresti, A. (2007), "An Introduction to Categorical Data Analysis", 2nd Edition, Wiley, ISBN: 978-0-471-22618-5, http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471226181,descCd-tableOfContents.html. See also http://www.ats.ucla.edu/STAT/examples/icda/default.htm

Tritchler, D. "An Algorithm for Exact Logistic Regression", Journal of the American Statistical Association, Vol. 79, No. 387 (Sep., 1984), pp. 709-711, http://links.jstor.org/sici?sici=0162-1459%28198409%2979%3A387%3C709%3AAAFELR%3E2.0.CO%3B2-%23

Tritchler, D. (1999), "Reasoning about data with directed graphs", Statistics in Medicine, Volume 18, Issue 16, Pages 2067 - 2076, http://www3.interscience.wiley.com/cgi-bin/abstract/63001837/ABSTRACT

Tritchler, D. (1999), "Modelling study quality in meta-analysis", Statistics in Medicine, Volume 18, Issue 16, Pages 2135 - 2145, http://www3.interscience.wiley.com/cgi-bin/abstract/63001838/ABSTRACT (PM: why not add the data ? Meta-analysis is a way to treat the Simpson paradox)

Yao,Q. and D. Tritchler (1993), "An Exact Analysis of Conditional Independence in Several 2 x 2 Contingency Tables", Biometrics, Vol. 49, No. 1 (Mar., 1993), pp. 233-236, http://links.jstor.org/sici?sici=0006-341X(199303)49%3A1%3C233%3AAEAOCI%3E2.0.CO%3B2-4